Active Learning in STEM
&
Biology Learning & Teaching in the Laboratory Context

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE UNIVERSITY OF MINNESOTA
BY

Ngawang Youdon Gonsar

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTORATE IN PHILOSOPHY

Dr. Bhaskar Upadhyay
Dr. Sehoya Cotner
(Advisors)

July 2021

Acknowledgments

The work detailed in my dissertation was possible through the guidance and support from my advisors, Professors Sehoya Cotner, and Bhaskar Upadhyay. I am thankful to Professor Cotner for helping me shape my dissertation and for supporting me throughout my doctoral journey. I am grateful to Professor Upadhyay for always encouraging me to expand my bounds as a scholar and for his constant support. I also extend my gratitude to Professor Sue Staats, for serving on my committee members and for nourishing my scholarly growth. I am ever so grateful to have previously trained under Professor Jennifer O. Liang. Thank you for also serving on my committee and always supporting my growth. I would also like to thank the Professors that I've had the pleasure of learning from over my doctoral journey. I want to also thank the Dalai Lama Trust Scholarship and the Community of Scholars Program for their support in my doctoral journey.

Chapter 3 Acknowledgments

We would like to thank Kendall Edstrom, Monica Osterbauer, Maryam Karaborni, and Ryan Laffin for their help in coding the qualitative responses of the participants. Thanks also to Michael Waltonen, Michael Balak, Michael White, John Ward, and Paul Strykowski for connecting us to their students. We appreciate the helpful input of #therealcotnerlab on an earlier version of this manuscript. A Cotner Lab Winter Writing retreat at Lake Itasca Biological Station and Laboratories was critical for initial data interpretation and discussion.

Chapter 4 Acknowledgments

We would also like to thank Maryam Karaboni, Kim Tran, Sergio Molina, and Alyssa Olson for their help in coding the qualitative responses of the participants. Finally thank you to members of the #therealcotnerlab for their input at various stages of this work.

Chapter 4 & 5 Acknowledgments

I would like to thank the students in the biology course for being willing to participate in this study and my dear colleague, Amy Kochsiek for her support throughout this project. This work would not be possible without them.

Dedication

I want to dedicate this work to my late Pala, Wangdu Phunstok Gonsar. I wish we could have spent a little more time together but alas it wasn’t to be. Thank you for demonstrating through example what it means to live a life of service and compassion. I hope to continue on that journey you set forth. I want to also extend my dedication to my dear family and friends. To my family, thank you for your love and support, always. Thank you to Gakyi. This would not have been possible without your constant support, every step of the way. My dear friend Brian, thank you for always being a sounding board for all my ideas and for always challenging me and thereby making me a better scholar.

Combined Abstract

This three-paper dissertation addresses the experience and the implementation of evidence-based learning practices in science, technology, engineering, and mathematics (STEM)/biology education. Study 1 explored instructional strategies and student perceptions and preferences for various teaching practices in graduate and undergraduate classrooms across three STEM colleges. The study revealed that students desired more time for active learning practices and wanted fewer lectures than they currently experienced. Upon closer probing, findings suggest that educators should employ various active learning practices in their classrooms. Finally, the study provides suggestions for instructors teetering on the brink of adoption to leap into active learning.

Study 2A and 2B narrowed the focus to learning in groups, which is the most utilized active learning strategy in biology courses. These studies examined how grouping strategies (self-selected vs instructor-assigned academically heterogeneous groups) impact first-year biology students' experience, performance, and cooperative learning participation in a biology laboratory course with extensive group work through a mixed-methods approach. There were similar effects on student perceptions from intervening in group strategies. However, students found substantial value in their group experiences in developing both academic and social skills. At the same time, students experienced diminishing concerns regarding their group members over time. When examining cooperative learning, there were many similarities but a greater frequency of cooperative learning elements when controlled for teacher's influence and the curriculum activity. There was also a small difference in the scores of assignments completed as a group.

These findings in totality have implications on how instructors can best form groups that maximize student learning while improving students’ biology laboratory experience. The study findings suggest that once pedagogical approach and curriculum are controlled, there is evidence that academically heterogeneous groups, as opposed to self-selected groups, allow for more cooperative learning opportunities for first-year biology students.

Abstract

Chapter 3

Despite positive evidence for active learning, lecturing dominates STEM higher education. Though most instructors acknowledge active learning is valuable, many resist implementing active learning techniques--citing an array of barriers, including a perceived lack of student buy-in. However, few studies have explored student perceptions of specific active learning teaching practices, particularly graduate students' perceptions. Chapter 3 explored student-reported instructional strategies and student perceptions of and preferences for various teaching practices in graduate and undergraduate classrooms across three STEM colleges at a large, public, research university. The study found that both graduate and undergraduate students desired more time for active learning and wanted less lecturing. However, there was no single universally desired or undesired teaching practice, suggesting that a variety of active learning teaching practices should be employed in both graduate and undergraduate courses.

Chapter 4

While there is consensus on the benefits of cooperative learning, there is less consensus on how best to form groups that maximize learning and improve the student learning experience. The conflicting results in group formation on maximizing learning and student perceptions for group work have led to the exploration of how undergraduate students experience working in different groups in a first-year biology laboratory course. Evaluating student perceptions and their experiences is essential because their perceptions affect their behaviors, response, and beliefs about the environment.

Therefore Study 2A explored how first-year students experienced extensive group work in a biology laboratory course via a mixed-methods approach using two different methods of group formation. First, student experiences while learning in self-selected groups, the usual method most common within the discipline, were explored. Next was an examination of student experiences with learning in instructor-assigned academically heterogeneous groups. There were similar effects on student perceptions from both grouping strategies. Study 2A indicated that students find substantial value in their experiences with group work in developing their academic and social skills. At the same time, students experienced diminishing concerns regarding group members over time. These findings have implications on how instructors can best form groups that maximize student learning while improving students' laboratory experience.

Chapter 5

Study 2B followed up on the previous findings by examining for any differences or alignments in academic performance for students in assigned and self-selected formed groups. Study 2B also examined students from the same study population in both the assigned and self-selected groups for evidence of cooperative learning using a validated observation protocol. There were many similarities between group types but there was a greater frequency of cooperative learning elements in the assigned group when controlled for the instructor and the lab curriculum. In addition to the difference in cooperative learning experiences from group type, there was also a small positive difference in the scores of assignments completed as a group. Findings suggest that for instructors keen on cooperative groups, the instructor's pedagogical approach may be more vital than the

grouping strategy. For instructors that are more aligned with cooperative learning, group type does not appear to matter as much. However, group type does appear to affect instructors who may be less familiar with cooperative learning.

Table of Contents

List of Figures

List of Tables

I. INTRODUCTION

Towards Discipline-Based Pedagogical Changes: A Historical Perspective

Science, technology, engineering, and math (STEM) teaching in the past decades: "It is a tradition. It was part of my training and seems like what I should be doing. I feel somehow guilty when I am not lecturing," (Creed, 1986, p. 25). Meta studies from the 1980s revealed that students experienced lectures, a teacher-centered approach, as the primary instruction mode during their undergraduate experience, including in their STEM classrooms. In courses more aligned with STEM, including 89% of physics and mathematics, faculty relied on lecturing as the sole means of teaching. Lecturing was and remained perceived as synonymous with teaching. It was and is the dominant method by which the faculty themselves were taught – it was the method by which most of them continued to teach (Bonwell and Eison, 1991; Akiha *et al*, 2018; Stains *et al*, 2018).

Discovery of issues in STEM education: Beginning in the mid-1980s, the work of the Higher Education Research Institute (HERI) highlighted many of the concerns plaguing STEM (then known as SME) higher education. These concerns included the drop in first-year students choosing STEM-based programs. A nationwide study of 2-year and 4-year colleges revealed a 40% loss of STEM majors, most of these taking place within the first two years into college (Hilton and Lee, 1988; Astin and Astin, 1993). Moreover, the number of students who joined STEM majors after their enrollment into college remained minimal.

The general drop in STEM majors also highlighted that STEM professions and the majors supplying them were disproportionately white and male (Seymour, 2002). The same HERI studies highlighted a 20-year decline in female enrollment despite increased recruitment efforts. Not only were enrollment rates lower, but the persistent rates of women in STEM majors were also lower than those of the men (Astin and Astin, 1993).

Similar trends were also ongoing in other underrepresented groups, including Hispanic, African American, and Native American students (Astin and Astin, 1993). Only a third of the Hispanic students, and one-half of African American and Native American students enrolled in STEM majors, graduated with a STEM major. National data demonstrated that only 38% of students of color entering engineering colleges graduated, compared to 68% of white students (Morrison and Williams, 1993). Likewise, between the first and third years of college, 65% of students of color entering science or mathematics dropped out of their majors (Culotta, 1992). Of the students who dropped out, half switched to non-STEM majors; half of the students who left engineering left college altogether (Campbell, 1993).

The under-representation concerns generated a debate about other issues, including the inequity in educational and occupational access, criticism of the quality of the STEM college experience, and finally, the effect on the workforce. Even with the pool of young white males exploited, the workforce demanded more STEM majors, forcing alternative talent sources to be tapped (Seymour, 2002).

Efforts towards improving STEM education: By the early 1990s, these concerns generated a national effort to recruit underrepresented groups into the sciences. The National Science Foundation (NSF) and the National Institute of Health (NIH) spent over a billion dollars to increase the participation of historically underrepresented students (Sims, 1992). These programs were effective in recruiting, but retention remained troubling. Around 2 out of every 3 Hispanic students and 1 out of every 2 African American students left science, math, and engineering majors (Astin and Astin, 1993). The lack of success from these early programs may have been because the programs were solely targeting underrepresented groups instead of focusing on the inequities of STEM education (Seymour and Hewitt, 1997). Additionally, programs targeting underrepresented groups deflected attention from the more significant challenge—to improve the quality of the undergraduate learning experience for all students (Seymour, 2002).

Science for all and pedagogical changes: Since the late nineties, the focus has been to improve the STEM competencies of all students, a mindset that a "*rising tide lifts all boats*"—improvements in the general quality of college STEM education will benefit all students, but will also disproportionately aid those who are poorly served by the existing undergraduate learning experience (Seymour, 2002). There is now a growing network of faculty experimenting with pedagogy in their classroom, departments, and networks for this effort. As a result, a shift towards *science for all* (AAAS, 1990) has had implications on teaching in the classroom, deemphasizing passive learning, and emphasizing active learning.

Organization of my Dissertation

Towards this effort, my three-paper dissertation explores science learning and teaching practices informed by constructivism, including active learning and cooperative learning in the STEM/Biology higher education context.

In Study 1 (Chapter 3), I examined the current teaching and learning practices prevalent across the STEM higher education curricula and identified ways to bring evidence-based teaching into classrooms. Study 1 is a single study that includes an introduction, methods, results, and a discussion section (Gonsar, Patrick, Cotner, 2021).

In Study 2 (Chapters 4 and 5), I shifted my focus to examine grouping strategies critical for successful cooperative groups. Through a mixed-methods approach, I explored grouping stages in a first-year biology laboratory course. I examined how students experience working in extended (semester-long) groups, its potential impacts on performance, and how it affects cooperative learning participation. Study 2 encompasses 2A and 2B of an overall project. Each part includes an introduction, methods, results, and a discussion section.

A final Chapter 6 includes my central findings from all three studies and their implications for learning and research in the future.

Chapters 1 and 2 serve to introduce and contextualize the three studies, respectively. In Chapter 1, I provide a brief roadmap for navigating my three-paper dissertation proposal. Chapter 1 also serves to provide a broad overview of the changes in the process for undergraduate STEM education. I include a historical perspective on STEM education trends and detail how my work contributes to the growing literature on Science learning and teaching practices, "*science for all*." Chapter 2 provides a framework for my three studies through an in-depth literature review and a deep dive into selected methodological approaches.

Research Design

Study 1: Graduate and Undergraduate-Student Perceptions of and Preference for Teaching Practices in STEM classrooms

Despite the evidence in support of active learning, reliance on lecturing remains pervasive in STEM higher education. Though many instructors acknowledge the value, they resist implementing active learning techniques in their courses. When queried about their reluctance, faculty often cite an array of barriers – most notably, a perceived lack of student buy-in for active learning practices. On the other hand, few studies have explored student perceptions of active learning or examined active learning practices in graduate-level courses.

Chapter 3 (Study 1) of this dissertation examines the student perceptions of 23 teaching strategies in graduate and undergraduate classrooms across the three STEM colleges at a large mid-western university. Based on the findings, I proposed key takeaways for instructors keen to implement evidence-based practices, including active learning methods in their courses. I identified that traditional lecturing was the most experienced mode of teaching in both graduate and undergraduate classes through a survey-based approach. Both graduate and undergraduate students overwhelmingly desired more time devoted to active learning than was experienced in their large STEM classes. Both these populations also wanted less lecturing as the primary mode of instruction. This study also demonstrated that no single active learning practice was universally preferred or unwanted. These findings suggest that instructors should implement various active learning practices in their classrooms, for graduate and undergraduate students alike. Finally, the results have implications for faculty professional-development programs at all levels of post-secondary instruction. This work is in the accepted manuscript: Gonsar, N., Patrick, L., Cotner, S., (2021). Graduate and Undergraduate-Student Perceptions of and Preference for Teaching Practices in STEM classrooms. Submitted on Dec 20th, 2020 *Disciplinary and Interdisciplinary Science Education Research (DISER).*

Our ***Research Questions*** for Study 1 were:

1. How do undergraduate and graduate students perceive the teaching practices in their curricula? Specifically, which teaching practices do undergraduate and graduate students experience, and which do they prefer?
2. Are there notable differences in undergraduate and graduate-level courses concerning the implementation and perceptions of active learning?

Study 2A: Cooperative Learning Groups in a Biology Laboratory Course: Exploring Grouping Strategies (2A)

Many studies cite the benefits of peer learning groups, but there is limited consensus on group forming for maximizing learning and improving students' group learning experience. Some recommend that heterogeneous groups are most effective for cooperative groups (Heller and Hollabaugh, 1991; Miller *et al*, 2012), and others found benefits with self-selected groups, common in the STEM disciplines and which are often homogeneous in gender, ethnicity, and or academic ability (Lou *et al*, 1996; Baer, 2003;

Jensen and Lawson, 2011).

Chapter 4 (Study2A) of this dissertation examined how first-year students experienced self-selected and instructor-assigned heterogeneous methods of group formation while engaged in extended group learning in a biology laboratory course. There were similar effects on student perceptions of cooperative group work from intervening in-group heterogeneity. The findings indicate that students find tangible value in developing academic and social skills in their group work experiences regardless of a group type. Students found their groups most benefited them by *contributing different perspectives* and bettering their *understanding* {of} *the material*. At the same time, student's concerns regarding working in groups diminished as they continued working together. In combination with previous work, conclusions from Study 2A recommend that instructors ensure more stable assigned peer groups. Also, students should form and experience more group work. This work encompasses a complete manuscript ready for submission.

My ***Research Questions*** for Study 2 were:

1) How is learning within assigned and self-selected base groups perceived by undergraduate students?
2) How did the perspectives of students change (if at all) after working in self-selected and assigned groups for a semester's duration?

Study 2B: Cooperative Learning Groups in a Biology Laboratory Course: Exploring Elements of Cooperative Learning and Academic Performance (2B)

Evidence in other learning spaces suggests academic and social skill benefits with cooperative learning (Slavin, 1980; Felder and Brent, 2007; Johnson and Johnson, 2018). However, there is limited knowledge on how students in the laboratory setting engage in cooperative learning elements and how grouping strategies influence their cooperative learning experience.

Chapter 5 of this dissertation examines cooperative learning in a first-year biology laboratory. Building on Part 2A, 2B examined for any differences or alignments in academic performance for students in assigned and self-selected groups by comparing students' current course grades against their grades in the preceding course, delineating any group-based and individual assignments differences. The same population of students from 2A was also examined in 2B for evidence of cooperative learning (CL). CL was observed as students engaged in extended group work in a biology laboratory course by employing the Cooperative Learning Observation Protocol (CLOP) (Kern *et al*, 2007), a validated observation protocol. There were many similarities between group types but a greater frequency of cooperative learning elements in the assigned group when controlled for the instructor and the lab curriculum. Students in assigned and self-selected groups primarily had similar academic performances. However, when grades were delineated, assigned groups secured gains in group work beyond the self-selected group suggesting a marginal positive difference for the assigned group. Study 2B's findings suggested that an instructor's pedagogical approach may play a critical role in the evidence for cooperative learning. In particular, the most influenced cooperative learning elements were group processing and promotive interaction. In combination with findings in 2A, the findings in 2B implies that students require structured group work with scaffolded instructional support paired with an effective curriculum for reaping the benefits of cooperative learning. Finally, there were more similarities in the evidence for cooperative learning regardless of group type when considering the aforementioned factors. This work encompasses the complete manuscript ready for submission: Gonsar, N., Pokorny, A., Cotner, S., (2021) Cooperative Learning Groups in a Biology Laboratory Course: *Exploring Elements of Cooperative Learning and Academic Performance (Study 2B)* Ready for submission 7.12.2021 *International Journal of STEM Education*

Our ***Research Questions*** for Study 2B are:

1) What, if any, are the elements of cooperative learning prevalent in self-selected and assigned groups in an introductory biology laboratory course?
2) How, if at all, does academic performance vary between students in the self-selected and assigned groups?

II. LITERATURE REVIEW

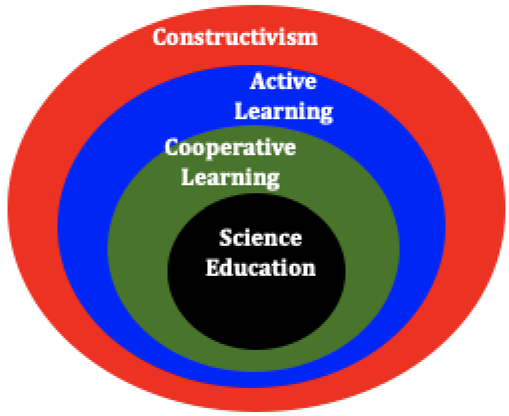


Figure 1: Illustrative contextualizing of Science Education using concentric circles. Science Education is an area of application of Cooperative Learning. Cooperative Learning is a form of Active Learning and Active Learning is informed by Constructivism.

Defining and Framing Constructivism in Higher Education

Constructivism views learning as an active process (Dewey, 1916). Social learning theory (socio-constructivism) is a variant of constructivism, drawing from philosophers like John Dewey. Below I focus on social learning theory as it concerns learning science in higher education. I specifically examine the works of Vygotsky for his social learning perspectives.

In a constructivist approach, learners internally build knowledge structures from experience, instruction, and prior knowledge foundations (Bransford *et al*, 1999). As such, social learning theories require learning to begin from a student's prior knowledge. For example, in an undergraduate learning space, instructors can set up pre-lecture questions or clicker questions to investigate students' prior knowledge on the topic at hand. Once prior knowledge is known, it is used as the foundation to design instruction from where students begin their learning process (Handelsmaan *et al*, 2004).

Building off Dewey's work, social learning theorists rejected the notion of students as "empty vessels" needing teachers to fill them with knowledge. Instead, they advocate for students to actively engage in their learning process (Bransford *et al*, 1999). In higher education, such forms of learning have manifested in active learning, a diversity of teaching methods aimed at engaging students in constructing knowledge. Some examples of active learning in higher education are think-pair-share (Tanner, 2013), classroom-response systems (e.g., "clickers"; Mayer *et al,* 2008), and one-minute papers; worksheets, completed by individual students or in groups; or collaborative group work through case-studies, problem-based learning or process-oriented guided inquiry learning (POGIL; Eberlein *et al*, 2008).

Teaching within a social learning theory approach to learning: Social learning theorists argue that learning through social interactions is essential for cognitive development. Therefore, learning is a social phenomenon. Recognizing learning as a social process was a significant shift from the previously dominant developmental theory (Tudge, 1993). Within a social context, social learning theorists, including Vygotsky, advocated that learning is in the interaction of teachers/capable peers with learners (Vygotsky, 1980). When learning is successful, knowledge is co-constructed, and the learner progresses towards cognitive development (Vygotsky, 1968).

To demonstrate the collaborative nature of learning, Vygotsky formulated the zone of proximal development (ZPD). The actual development is the ZPD stage, already mastered by a student, serving as the inner boundary of the ZPD. A student moves towards the potential development stage with the assistance of teachers or through collaboration with more capable peers. Once students reach the potential development stage, they encounter the outer boundary of the ZPD. Beyond this stage, the student has grasped the concepts by internally building cognitive structures leading to the co-construction of knowledge. In various learning spaces, including undergraduate education, scaffolding is a common strategy implemented to help learners move across their ZPD. For example, in an undergraduate interdisciplinary research methods course, scaffolding was successfully implemented to expose students to various scientific disciplines and become familiarized with their research methods and design. As a result, students by the end of the course had developed skills to ask their research questions and to put together a research proposal, where previously they were unable (Keebaugh *et al*, 2009).

The apprentice model is the ZPD in action and has manifested in apprentice research experiences (AREs) and the growing and more available course-based undergraduate research experiences (CUREs) in the undergraduate space. Apprentice models relate to small learning communities with specialized roles aimed at accomplishing goals that relate both to the group and the broader community (Rogoff, 1990). As such, the CUREs model focuses on the nature of the activity and the broader relevance and meaning of the activities beyond the scientific community. For instance, in a typical CUREs course, students initially model skills from the instructor or TAs but become fully participating members of a novel project over time. Thus, students carry out tasks independently and use scientific practices to engage and discover broad-relevance projects (Corwin *et al*, 2014).

Social learning perspectives include teaching practices encompassed under the active learning umbrella (Figure 2). Teaching practices are student-centered, where learners construct or co-construct their knowledge. Social learning theory emphasizes the learner's social context, an area where educators can intervene and develop performance metrics. Approaches grounded in social learning theory offer a road map with discernable consequences - reflected in the abundance of evidence-based research in this area. As a result, social learning theories have framed evidence-based teaching practices in higher education.

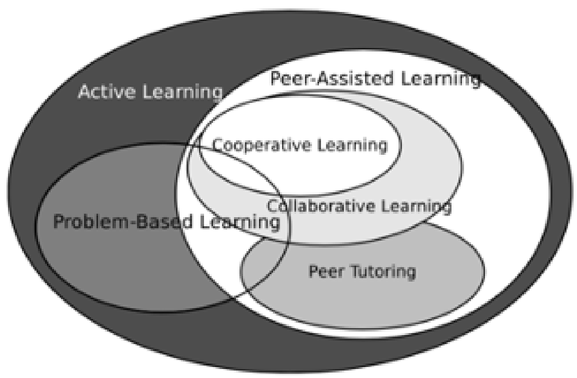


Figure 2: Venn diagram of learning theories and active learning methods (Bishop and Verleger, 2013).

Active Learning: Teaching and Learning in Post-Secondary Education

By definition, active learning is, "*instructional activities involving students in doing things and thinking about what they are doing,*" (Bonwell and Eison, 1991). At one end of the active learning spectrum, it can be merely pausing lectures to allow students to clarify and organize their ideas through discussion with a neighbor. The other end of the spectrum can encompass more complex activities, including using case studies as a focal point for decision making (Brame, 2016).

In the higher education context specifically, "active learning" is a term encompassing a diverse assortment of teaching practices in which students engage actively with the course content, instructor, and each other using various activities. Additionally, active learning practices are characterized by students involved in solving problems, reading, writing, and discussing (Prince, 2004). Overall, such methods have a greater emphasis on students' explorations of their attitudes and values than traditional ways of teaching (Bonwell and Eison, 1991). Specifically, some teaching practices with these attributes include group discussions, clicker questions, debates, and projects (Miller and Metz, 2014). These practices aim to involve students in higher-order thinking tasks, which lead to knowledge construction.

Cooperative Learning: Under the Active Learning Umbrella

In undergraduate science, cooperative learning, under the umbrella of active learning, is an existing learning strategy that stems from social learning theory (Figure 2). Learning methods centered around group work are critical for social learning theorists and have demonstrated increased learning gains in the science undergraduate space (Johnson and Johnson, 1991; Patrick, 2013).

Cooperative learning: Undergraduate science courses often have at least one group project during a semester (Tanner *et al*, 2003). The students in these courses come with diverse experiences and from a variety of backgrounds. These include students from high schools across the country and a growing number from outside the United States

(Singer and Smith, 2013). In addition, these groups of students possess varying levels of content knowledge and experiences. Collectively, these attributes make cooperative learning a valuable opportunity in science courses in both lab and lecture.

There is also increasing evidence that activities centered around cooperative learning increase student learning (Johnson and Johnson, 1991; Beichner *et al*, 1999; Ebert-May *et al*, 2003). For example, two groups learned about chemical bonding in an undergraduate chemistry course through cooperative learning activities and the traditional method, respectively. The study concluded that students engaged in cooperative learning techniques had increased understanding and outperformed students taught in the traditional methods (Karacop and Doymus, 2013). Moreover, as science undergraduate students progress within their disciplines and push forward lines of inquiry, they encounter increased needs to collaborate and access the expertise of various professionals from a diverse range of backgrounds, making skills garnered from cooperative learning vital (Tanner *et al*, 2003).

Cooperative learning involves purposely planned small groups of students working towards a shared learning goal (Deutsch, 1949a; Bybee *et al*, 2010). For example, one study selected groups of three academically heterogeneous students for a physics course for the duration of a semester. The instructor assigned a specific role to each group member, rotating the positions for the next assignment to complete various conceptual and problem-based assessments (Beichner, 1999). The careful selection of the members aimed to create sustaining group members with interdependent attributes and with each member tasked to a specific role (Johnson and Johnson, 1991). In cooperative learning, once groups are successfully selected, the members can form a positive interdependence, invested in each other. As an outcome, each student is more motivated to make an equal contribution to the completion of the academic goal. An instructional design that promotes positive interdependence is by assigning a component of the overall grade from the group's performance (Johnson and Johnson, 1991). In cooperative learning, students also build social skills to function effectively and efficiently as a group.

Additionally, though knowledge is co-constructed with cooperative learning, individual students are responsible for demonstrating their understanding, as assessed by the teacher. For instance, in a cooperative learning project within a neurobiology class, students were evaluated periodically through individually written quizzes. A similar examination was then completed in learning groups, providing insights into the knowledge gaps of individuals within a group (Gaudet *et al*, 2010). Finally, the teacher in a cooperative learning class can form group members, each of whom possesses various attributes that strengthen the group's ability to meet the learning goals. A heterogeneous academic group is a way to heighten students' learning previously familiar and unfamiliar with the topic (Rewey *et al*, 1992).

Teaching a unit on introducing evolution by natural selection through cooperative learning through a social learning approach

Gauging prior knowledge and group selection: One of the fundamental tenets of a cooperative learning environment demands that groups sustain and contain students of interdependent attributes. Therefore cooperative groups can be formed towards the beginning of the semester. One way to create groups is via students of varying levels of understanding to be in the same group, forming heterogeneous groups (Johnson and

Johnson, 1991; McInergney and Fink, 2003; Gaudet *et al*, 2010). A metric to form heterogeneous groups is through the implementation of a pre-assessment.

The pre-assessment would broadly cover various topics students learn across a semester to determine their current understanding (cell structure and function, gene structure and expression, evolutionary processes). Finally, if an instructor teaches a continued part of a course (i.e., General Biology 1 & 2), the students' performances in the previous course must be considered. Based on these criteria, the instructor should select groups of 2-3 students.

Illustrative example of Cooperative Learning Manifestation in the class: To cement their understanding of natural selection, students conduct an exercise. The activity tasks group members as the predator, helper (setting up the prey), and the timer. Once roles are assigned, each group engages in a simulation game with the same prey of different phenotypes but with various heterogeneous environments. The prey is, of course, subject to natural selection by the introduction of the predators. Conditions of no immigration/migration/spontaneous mutation are specified ahead of time. The activity is time-based, encouraging students to coordinate and communicate effectively with each other.

Predation occurs every generation, with group members switching roles for each event. The assigned roles remain procedural and not based on student ability. Once multiple rounds of predation and reproduction of prey have occurred, students work together on structured questions assigned by the instructor. A structured question is a form of scaffolding technique beneficial in guiding students towards their potential development.

The students work together on completing the questions. The questions should deduce student understanding about natural selection as derived from the activity. By doing so, students are invariably co-constructing knowledge by drawing their knowledge from the simulation to their overall understanding of natural selection. In addition, engaging in the activity and working together on the questions allows the student to build social skills and work cohesively as a group.

Before the activity starts, instructors inform the students that they will receive a collective grade for the group. A collective group grade is essential to increase the interdependence within the cooperative group. Additionally, instructors can also gauge each student's understanding by following up with an individual quiz. If students successfully constructed their understanding of natural selection, they will have reached the outer bounds of the ZPD by working as a group.

Teaching informed by social learning theory will use scaffolding techniques to help students gain mastery and for the learner to discern implications of their co-constructed knowledge in the broader community (Sanders and Welk, 2005). To accomplish this, an instructor may share a video on antibiotic resistance, including its manifestation and effects in the world. The instructor will then pose an abstract question requiring the group to demonstrate how antibiotic-resistant bacteria arise. Next, students in the same groups explore the question. Finally, while working together as a group, students utilize their understanding of natural selection and their prior knowledge of prokaryotes to arrive at an explanation.

In the meantime, teachers are walking around each group. When needed, teachers use other scaffolding techniques such as prompts or questions. According to social learning theory, the interactions between peers, learners, and instructors are happening within the ZPD. Therefore learning is taking place. Once students have gained mastery, they can demonstrate their understanding through problem-solving skills (Vygotsky, 1980). To test for this, an instructor can verbally ask a problem-solving question such as, "*Create a proposal as a researcher to reduce bacterial strains resistant to the drug?*" Being able to demonstrate problem-solving skills demonstrates mastery of the topic.

Study 2 Approaches and Methodology: A Deeper Dive

Survey Items Validation via the Think-Aloud Protocol

Think-aloud is a type of validity testing (Charters, 2003). It asks respondents to read through an instrument such as a survey and discuss out loud how they interpret the questions and how they would arrive at their responses. Survey instruments can provide a reliable response. The questionnaires in a survey can provide numeric and textual directions of trends, attitudes, and opinions of the population by studying a population sample (Creswell, 2009). From the sample results, we can describe trends in attitudes and behaviors by gathering data from the sample, analyzing the data to identify average opinions, the prevalence of behaviors, or the range of individuals' attitudes. By doing so, we are making certain generalizations about the population at large. However, without validating, participants can respond to a survey but can interpret the survey items differently from those who designed the study.

The think-aloud technique requires research participants to verbalize their thoughts and perceptions while they engage with the survey (Charters, 2003). As the lead researcher, I probed the participants on their reasoning for selecting a particular response while filling out the survey items. As participants detailed their reasoning, I wrote down precisely what was said by each of the participants. The think-aloud technique is particularly useful in determining whether participants interpret the survey questionnaires the same way intended by the researchers.

While conducting a think-aloud technique, I listened to how the participants think about the questionnaire's instructions and the survey items. I paid particular attention to whether any of the questionnaire items were confusing or threatening. I asked participants to tell me when they reached an item that was difficult to understand and then asked them to paraphrase what they believed the problem item was stating or asking. I made sure all the participants understood the survey items in the same way.

Early research recognized that five participants could detect 80% of survey usability problems (Virzi, 1992). Of the five, the first three will discern most of the usability problems, and beyond that, there are gradual diminishing returns. To meet this criterion, I had five participants and ensured that the participants selected in the think-aloud sessions are similar to the survey participants.

Below is a summary of the think-aloud protocol steps I utilized for the pre and post-semester survey (Willis *et al*, 1991; Anderson, 2004; Stone *et al*, 2011).

- Informed the participants that the purpose of the think-aloud protocol is to evaluate the survey aloud, not just complete it.
- Probed participants as they thought aloud about each item. I asked questions of the participants that include:

1. What do you think this question is asking you?
2. How do you think you would answer the question?

3. Is this question confusing?
4. What does this word/concept mean to you?
5. Is there a different way you would prefer to respond?

- Asked participants 3 reflective questions as they reached the end of the survey. They included, a) Looking back, does anything seem confusing? b) Is there something in particular you hoped was going to be asked but wasn't? c) Is there anything else you feel I should know to entirely understand this topic?

Qualitative Coding

Though approaches differ across researchers, all data that are analyzed qualitatively follow the same essential steps. These include coding the data, combining the codes into broader categories, and representing the data in figures and tables, incorporated in a discussion (Creswell, 2018). Qualitative coding analyzes and interprets textual data within a study (Miles and Huberman, 1994).

First cycle coding: For the first cycle of coding, I used an inductive *in vivo* method, generating coding from the words or phrases used by the participants (Strauss, 1987). In vivo coding is particularly useful for studies that prioritize and honor the participants' voices. It is a form of initial coding employed commonly during a grounded theory approach but can be used for qualitative coding in general. Therefore, I used *in vivo* coding as the sole coding method for the first data analysis cycle.

Reading through the transcript, I attuned myself to words and phrases that call for bolding, underlining, italicizing, highlighting, or vocal emphasis. If the exact words or phrases or variations thereof showed up frequently, it likely merited a code. I stuck to the traditional approach of generating *in vivo* (or other) codes via a line-by-line approach.

The goal of developing *in vivo* codes is to, "*help us (researchers) preserve participants' meanings of their views and actions in the coding itself,*" (Charmaz, 2006, pg 55).

Second cycle coding: Upon completing the first coding method via *in vivo* coding, I implemented the second cycle of coding methods. The second cycle of coding is a way of organizing and re-analyzing data coded through the first cycle method, requiring "*...linking of seemingly unrelated facts logically, of fitting categories one with another to develop coherent meta-synthesis of the data at hand,*" (Morse, 1994). Therefore, I developed categories from the array of the first cycle codes through the second cycle coding.

Steps for second cycle coding: First cycle codes are reorganized and reconfigured to eventually develop a smaller and more select list of broader categories, themes, and constructs (Saldana, 2009). For my study, I implemented pattern coding as my second cycle coding method. Using the pattern coding, I developed categories identified from the open responses (Miles and Huberman, 1994).

To code the open-ended responses from our study, I read and then reread all the open-ended responses (Agar, 1980). The following are three examples of open-ended questions from our research that required qualitative coding:

1. Please explain your overall perceptions of the utility of group work.
2. Please explain your typical approach to group work.
3. Please specify any concerns you have about working in a group.

In summary, I reviewed the purpose of the survey questions and what I want to find out from the responses. I implemented two cycles of inductive codings in vivo for the first cycle and pattern coding in the second cycle to best explore the responses. First, I read through the text and identified issues that recur in the data. These became our codes. This approach allowed codes to emerge from the data. Next, I organized the data into codes for every question while examining patterns and connections within and between the questionnaire codes. I then highlighted variations and assessed the areas where categories differed. I also identified the key ideas expressed within the codes and the categories. Finally, I determined how participants responded similarly to each other and how their responses varied. To recognize which categories are most common across responses, I counted the number of times particular codes and categories come up. These counts provided an estimate of relative importance, revealing the general patterns in the data.

Research Approach: Mixed Methods

Mixed methods allow for a deeper understanding of the problem/issue than a single method study allows for alone (Morse, 2003). Mixed methods were first implemented in 1959 by Campbell and Fisk in studying data validity in psychological traits. An initial rationale for utilizing mixed methods was in recognizing the limitations of all methods. Furthermore, mixed methods came to be viewed as a neutralizing factor to the biases of a single method (Jick 1979). Since then, other disciplines, including educational research, have employed mixed methods. In academic research, mixed methods have primarily used field method approaches, such as interviews and observations combined with traditional surveys (Sieber, 1973).

There are currently three main strategies for mixed methods. These include sequential, concurrent, and transformative mixed methods. Sequential mixed methods allow a researcher to expand on the findings of one method with another. There are variations within this strategy that enable the research to begin with a qualitative approach and generalize it with a quantitative approach or vice-versa to understand the data collected.

A convergent mixed-method is ideal for Study 2. A convergent mixed method allows a researcher to simultaneously collect both data sets (qualitative and quantitative) to provide a holistic understanding of the research problem (Figure 3). With this strategy, the qualitative data can be embedded within a quantitative study or vice versa, depending on the research questions (Clark and Creswell, 2014).

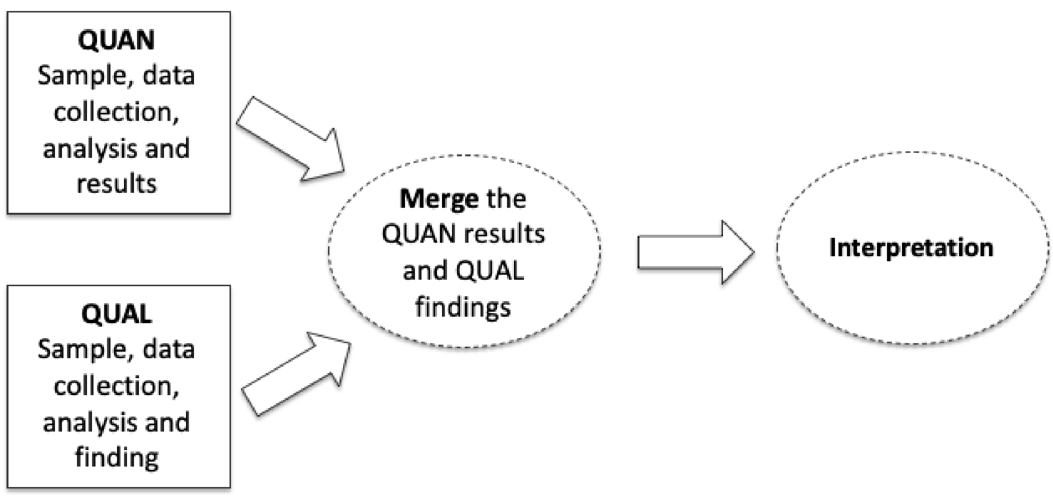


Figure 3: Illustration of a convergent mixed-methods research approach. The proposed study entails an equal contribution of qualitative and quantitative data to explore the research question(s) (Clark and Creswell, 2014).

III. GRADUATE AND UNDERGRADUATE STUDENT PERCEPTIONS OF AND PREFERENCES FOR TEACHING IN STEM CLASSROOMS

Introduction

STEM higher education is undergoing rapid change, driven by an increase in the number and diversity of students, digitalization and globalization, and shifting demands from policymakers and society at large (Shin and Harman, 2009; Brewer and Smith, 2011; Olson and Riordan, 2012; Graham *et al*, 2013). Simultaneously, research into evidence-based pedagogy has revealed that traditional, lecture-based teaching is not only ineffective overall, but disproportionately disadvantages women, first-generation students, and students from underrepresented groups (Haak *et al*, 2008; Ballen *et al*, 2017; Theobald *et al*, 2020). As a result, instructors are encouraged to teach using evidence-based approaches that increase student motivation, collaboration, and metacognition, all of which influence students' learning and course performance in STEM (Council, 2003; Glynn *et al*, 2011; Tanner, 2013a). These challenges and expectations directly impact instructors, who may lack the time, funds, and extrinsic motivators to think deeply and scientifically about teaching (Miller and Metz, 2014; Gormally *et al*, 2016; Patrick *et al*, 2016).

Evidence-based teaching is an umbrella term that includes active learning and other teaching practices shown to positively impact student learning (Felder *et al*, 2000; Owens *et al*, 2017). Active learning is itself a catch-all phrase, derived from constructivism, a learning theory that proposes that students learn by constructing their own knowledge (Freeman *et al*, 2014). Within a constructivist framework, learning is an active process and builds on experience, instruction, and the foundations of prior knowledge (Bransford and Schwartz, 1999; Prince, 2004). In practice, active learning can include small-group discussion (Tanner, 2013b), classroom-response systems (e.g., “clickers”; Cotner *et al*, 2008), one-minute papers and worksheets, completed individually or in groups, and collaborative group work (e.g., via case-studies, problem-based learning, or process-oriented guided inquiry learning [POGIL] (Eberlein *et al*, 2008). Through engagement in active learning practices, students remain an integral part of the learning process by building meaning and constructing knowledge (Prince, 2004).

Despite the evidence in support of active learning (Freeman *et al*, 2014), lecturing remains a pervasive feature of STEM teaching (Akiha *et al*, 2018; Stains *et al*, 2018). Some faculty may choose to lecture because they are not convinced that active learning is effective (Silverthorn *et al*, 2006; Michael, 2007). Other instructors value active learning (Patrick *et al*, 2016), but refrain from integrating active learning techniques, citing an array of barriers including the time needed to prepare “activities” (Brownell and Tanner, 2012), lack of training in effective teaching techniques, lack of time for content coverage (or, the loss of lecture time), perceived lack of student buy-in (Cavanagh *et al*, 2016; Owens *et al*, 2017; Deslauriers *et al*, 2019), or the concern that their classes are prohibitively large for active learning (Silverthorn *et al*, 2006, Patrick *et al*, 2016). In this work, we focus on one of these perceived barriers—a lack of student buy-in to active learning pedagogies.

A significant predictor of active learning implementation and engagement with teaching practices is the faculty and students' buy-in, (Cavanagh *et al*, 2016; Madson *et* *al*, 2017). Faculty often fear student resistance to active learning (Silverthorn *et al*, 2006; Seidel and Tanner, 2013), impacting the classroom environment and their evaluations by students (Henderson *et al*, 2018). Although these evaluations are flawed (Uttl *et al*, 2017; Carpenter *et al*, 2020; Stroebe, 2020; Wang and Williamson, 2020) they remain meaningful to the faculty for merit pay, promotion, and tenure. We also recognize that student *preferences* do not always mirror the practices that lead to the most learning gains (Deslauriers *et al*, 2019) which is also reflected by work at our own institution (Cotner *et al*, 2008; Walker *et al*, 2008). Student resistance can be lowered with evidence, and previous studies have examined student perceptions of active learning in individual undergraduates (Bransford and Schwartz, 1999; Machemer and Crawford, 2007; Smith and Cardaciotto, 2011; Brazeal *et al*, 2016; Cavanagh *et al*, 2016, 2018; Patrick *et al*, 2016; Brown *et al*, 2017; Cooper *et al*, 2017; England *et al*, 2017; Owens *et al*, 2017; Brigati, 2018; Mcmillan *et al*, 2018) or graduate (Lopez and Gross, 2008; Jones *et al*, 2010; Tune *et al*, 2013a; Miller and Metz, 2014) courses. Although the opinions of individual students may differ, as a whole, both graduate and undergraduate students in individual STEM courses reported neutral through very positive perceptions of and preference for active learning teaching practices (Patrick, 2020). These studies provide valuable insight into how students view these teaching practices. However, these were studies of specific courses, and active learning implementation was controlled or its practice known to the researchers. As a result, it remains unknown if the findings represent student perceptions and preferences within a broader context.

Few studies have examined student perceptions of and preferences for teaching practices across STEM disciplines, where active learning remains limited, and student resistance is often a perceived barrier (Patrick *et al*, 2016, 2018). Patrick *et al*, (2016) and Patrick *et al*, (2018) examined student perceptions of active learning and other teaching practices among science college departments of a large research-intensive university in the southeastern United States. Using a modified survey (Miller and Metz, 2014), students were asked to estimate the amount of their science class time devoted to active learning and state the amount of time students thought should be dedicated to active learning. These studies also prompted students to rank six broad teaching practice categories most effective for their learning. Compared to graduate students, undergraduates reported less active learning in their classes. Nevertheless, both groups wanted more active learning than currently experienced (Patrick *et al*, 2016, 2018, Patrick, 2020). These studies used broad categories of teaching practices, making it impossible to interpret student perceptions of and preferences for specific teaching practices like think-pair-shares (Kaddoura, 2013). However, we can leverage student attitudes of particular teaching practices to increase faculty willingness towards such activities. Also, gauging the perceptions of different student populations is essential to learn how active learning and other teaching practices can be generalized in different contexts (Patrick, 2020).

As we have stated above, similar works in different higher educational contexts have not explored student attitudes towards particular teaching practices. To address this gap in knowledge, we surveyed students in three STEM-focused colleges at one large university in the midwestern United States to determine their experiences and perceptions of specific teaching practices. We compared student perceptions of the teaching strategies employed in their undergraduate and graduate-level courses to detect whether students valued different pedagogies at different stages of their education (i.e., undergraduate or graduate). We also compared the alignment between experienced and desired teaching practices in both undergraduate and graduate-level courses. Through our study, faculty and other stakeholders can be more fully informed and understand the instructional choices and student preferences throughout the STEM curriculum.

Research Questions

The main questions guiding this work were:

1) How do undergraduate and graduate students perceive the teaching practices in their curricula? Specifically, which teaching practices do undergraduate and graduate students experience and which do they prefer?
2) Are there notable differences in undergraduate and graduate perceptions regarding the implementation of active learning in their courses?

Materials and Methods

Institution and Study Participants

Our institution is a large land-grant university in the Midwest, serving 32,000 undergraduate and 16,000 graduate students. There are three central STEM Colleges: the College of Biological Sciences (CBS), the College of Science and Engineering (CSE), and the College of Food, Agriculture, and Natural Resource Sciences (CFANS). There is a total of 27 departments in these Colleges, with an aggregate enrollment of 12,096 students at the time of survey distribution. We were interested in comparing STEM undergraduate and graduate student experiences with active learning. Accordingly, any student enrolled in a degree program in either of the three STEM Colleges was in our target population.

Survey Instrument

Because we were interested in student perceptions of and preferences for specific teaching practices and active learning in general, we combined items from several existing survey instruments (Miller and Metz, 2014; Patrick *et al*, 2016; DeMonbrun *et al*, 2017). We asked all students to identify the largest STEM course they had taken in the preceding semester and to estimate the number of students enrolled. We asked this to encourage students to think about a single large-enrollment course and make the response sets more similar in the types of courses evaluated. For undergraduates, such large courses are often considered “gateways” to STEM majors and have received considerable attention from discipline-based education researchers (Barr *et al*, 2008; Xie *et al*, 2015; Witherspoon *et al*, 2019). For graduate students, a question on their largest course is likely to prevent them from considering a seminar or dissertation-credit course in their responses. For the identified course, we asked students how often the course instructor used specific teaching practices (Table 1). Most of the teaching practices included in the instrument were taken directly from the Student Response to Instructional Practices (StRIP) instrument (DeMonbrun *et al*, 2017). We modified two practices for clarity and added three known teaching practices at our institution (Table 1). For each teaching practice, students responded to the prompt: “*Please indicate how often each activity was done in the largest science course you took this semester*.” Frequency options were: Never or almost never (0-10% of the time) (scored as 1); Seldom (11- 30% of the time) (scored as 2); Sometimes (31-50% of the time) (scored as 3); Often (51-70% of the time) (scored as 4); Very often (71-100% of the time) (scored as 5). These options differed from those in the original StRIP to better align with the question about how much class time they think is and should be devoted to active learning included in our study. For each teaching practice, students were asked, "*How often would you like to do each activity in an ideal course you would take as a student?*" Response options were: Much less (scored as 1); Slightly less (scored as 2); About the same (scored as 3); Slightly more (scored as 4); Much more (scored as 5).

We also provided the students with a definition of active learning (Miller and Metz, 2014). After reading the definition, students estimated the percentage of class time typically devoted to active learning and how much time they think *should* be devoted via an open-ended response. Students also reflected on their experiences with active learning: "*Please describe your experiences with active learning in the classroom.*" Finally, we asked students to report their status (graduate or undergraduate).

Our study design and survey instrument were approved by our institution's IRB (approval #STUDY00002261).

Table 1. Teaching practices included in the survey instrument and their source.

Question number	Teaching practice	Source
Q1	Listen to the instructor lecture during class	StRIP
Q2	Brainstorm different possible solutions to a given problem	StRIP
Q3	Find additional information not provided by the instructor to complete assignments	StRIP
Q4	Work in assigned groups to complete homework or other projects	StRIP

Q5	Make individual presentations to the class	StRIP
Q6	Be graded on class participation	StRIP
Q7	Study course content with classmates outside of class	StRIP
Q8	Assume responsibility for learning material on own	StRIP
Q9	Discuss concepts with classmates during class	StRIP
Q10	Get most of the information needed to solve the homework directly from the instructor	StRIP
Q11	Be graded based on the performance of a group	StRIP
Q12	Preview concepts before class by reading, watching videos, etc.	StRIP
Q13	Solve problems in a group during class	StRIP
Q14	Solve problems individually during class	StRIP
Q15	Verbally answer questions posed by the instructor during class	Modified from StRIP
Q16	Verbally answer questions posed by the instructor during class after consulting with a classmate (think-pair-share)	Current work
Q17	Answer questions posed by the instructor during class using a student response system (clickers, TopHat, etc)	Current work
Q18	Answer questions posed by the instructor during class using a student response system (clickers, TopHat, etc) after consulting with a classmate (think-pair-share)	Current work
Q19	Ask the instructor questions during class	StRIP
Q20	Take initiative for identifying what is necessary to know	Modified from StRIP
Q21	Watch the instructor demonstrate how to solve problems	StRIP
Q22	Solve problems that have more than one correct answer	StRIP
Q23	Do hands-on group activities during class	StRIP

Note. Students were asked how often each teaching practice occurred in their largest STEM course and how often they would like each teaching practice to occur.

Survey Distribution

During the Spring 2018 semester, we contacted college administrators to obtain lists of current graduate and undergraduate students enrolled in the STEM colleges. We used the Qualtrics platform to distribute the survey and collect responses. We offered the first 100 respondents a $5 coffee card as an incentive to complete the survey. The survey was open for a total of two weeks. Two reminders were sent to students who had not yet completed the survey–one week and one day prior to the close of the survey.

Data Analysis — Quantitative

Responses were downloaded to Microsoft Excel from Qualtrics and de-identified by a researcher not otherwise affiliated with this project. We removed incomplete de-identified responses or those reported by individuals under the age of 18. Graduate and undergraduate student responses were analyzed separately. Significant differences between graduate and undergraduate students for activities desired and experienced were determined using the Mann-Whitney U Test. The Mann-Whitney test is most appropriate for ordinal data that deviates from a normal distribution (MacFarland *et al*, 2016). All data were analyzed using *Sigma plot 14.0*, the ggplot2 package for R, and RAWGraphs (Team, 2013; Wickham, 2016; Mauri *et al*, 2017).

Data Analysis — Qualitative

Seven hundred sixty-eight students responded to the open-ended prompt, "*Please describe your experiences with active learning in the classroom.*" Two coders, trained in *invivo* coding (Saldaña, 2009), read student responses, unaware of whether a graduate- or undergraduate-level student wrote the response. In an initial meeting, the coders identified consensus categories; afterward, they identified the following emergent categories (along with associate sub-categories) from the student responses: Positive About Active Learning (active learning a) makes the class engaging, b) is beneficial, c) helps with content retention, d) builds community and e) prepares for real-world work environment); Negative About Active Learning (active learning a) limits individual thinking and learning, b) professor does not implement active learning approach correctly, c) should only be used in particular fields and d) current active-learning methods are a "waste of time"); and Constructive (active learning a) works when people are prepared to collaborate b) only works in smaller classes, c) effective active learning is desirable, and d) works well when supplemented with other methods. Upon the generation of a sub-categories codebook, each coder coded the same randomly selected 10% of the comments to establish interrater reliability. Cohen's Kappa (κ) is a robust statistical approach for testing reliability while accounting for chance agreement between two raters. The raters received a κ =.87, considered a *strong* agreement on the Kappa scale (Cohen, 1960). Once reliability was established, we divided the remaining responses among the two coders. When the coding was completed, the sub-categories were tallied, and the responses were decoded to allow for comparison between graduate and undergraduate student responses.

Results and Interpretation

Participant Attributes

In total, 1274 undergraduate (n=1113) and graduate (n=161) students completed the survey (Table 2). Nearly equal numbers of first-year, second-year, third-year, and senior undergraduates responded, which together greatly outnumbered the graduate respondents. Respondents who identified as female outnumbered respondents who identified as male while 14 stated Other and 136 students opted to skip the question. The mean class sizes reported by graduate and undergraduate students were 39 and 178 students, respectively, for their largest science course.

Table 2. Student attributes.

Student status	**n**
Total	1274
Graduate	161
Undergraduate	1113
Freshman	268
Sophomore	259
Junior	272
Senior	247
Other	67
Gender	*n*
Female	697
Male	427
Other	14
No Answer	136

Teaching and Learning Practices Experienced by Students

Of the 23 teaching practices included in our survey, all students identified instructor lecturing (Q1) as the most common teaching practice in their courses, which occurred, on average, *very often* (Fig. 4A). This finding is consistent with other studies, which demonstrate that STEM classrooms are dominated by teacher-centered pedagogy with lecturing as the primary mode of instruction (Akiha *et al*, 2018; Stains *et al*, 2018). Graduate and undergraduate students also identified four other practices that, on average, happened *often* in their largest STEM courses: assuming responsibility for learning the material (Q8), getting homework information from the instructor (Q10), taking the initiative in deciding what is necessary to know (Q20), and watching the instructor demonstrate how to solve problems (Q21; Fig. 4A). These results suggest that all students experienced teaching and learning practices that were dominated by direct faculty-to-student, teacher-centered instruction.

A. B.

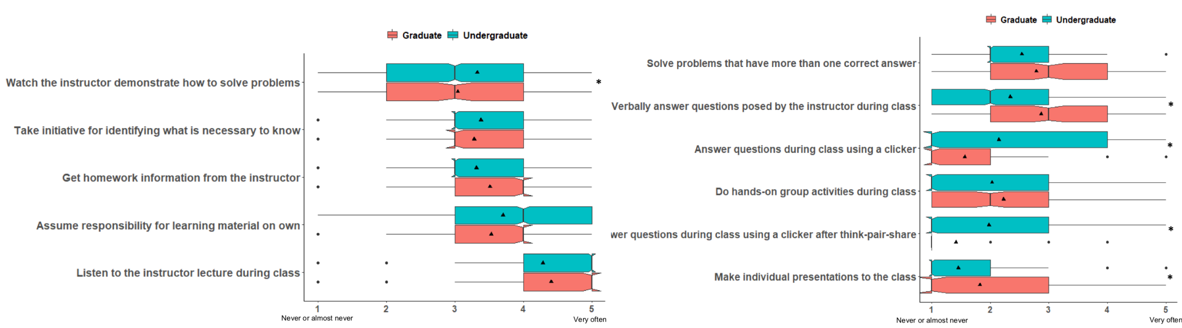


Figure 4: Box and whisker plots comparison of the most and least common instructional practices reported by graduate (*n*=161) and undergraduate (*n*=1113) students in STEM classrooms. (A) Most frequently experienced activities, and (B) Least frequently experienced activities. Triangles indicate mean values. *indicates p-values ≤ .05. Activities displayed are the five most and least desired. A reported sixth activity in panel B reflects the ranking misalignment between graduate and undergraduate students.

The teaching practices experienced by students *least often* were all active learning techniques, but these differed between undergraduate and graduate students. Undergraduates identified individual student presentations (Q5) as the least performed instructional practice (Fig. 4B). This finding is not surprising due to time constraints in large-enrollment courses. Teaching activities that entail direct feedback from instructors were also infrequently experienced. For example, students seldom experienced answering questions using student response systems—either directly or following a consultation with a classmate (Q17 and Q18; Fig. 4B). However, studies have demonstrated that students using student response systems in large classrooms are more engaged than those that do not use clickers (e.g., Mayer *et al*, 2008). Moreover, students retain more material on exam units covered in lessons that incorporate clicker activities (Crossgrove and Curran, 2008). Hands-on group activities and answering questions posed by the instructor were also uncommon practices in undergraduate courses (Q23 and Q15; Fig. 4B).

Graduate students reported the use of student response systems (i.e. Top Hat, clickers), both in isolation or following consultation with a classmate (think-pair-share), as the least common activities experienced in their courses (Q17 and Q18; Fig. 4B). We suspect that graduate-level instructors may be less inclined to use a classroom response system given the smaller class sizes of such courses. Giving individual presentations (Q5), answering questions verbally in the classroom (Q14), and solving problems that have more than one correct answer (Q22) were also infrequent in graduate courses (Fig. 4B).

Our findings for how often all 23 teaching practices were experienced by students in our sample can be found in the lower panel of Fig. S1.

Teaching Practices Preferred by Students

Undergraduate and graduate students desire significantly more class time for active learning pedagogies than they are experiencing (Fig. 5). On average, undergraduate students reported 31% of class time was *currently* being devoted to active learning and that 36% of class time *should be* devoted to active learning. Graduate students reported that significantly less class time, 25%, was *currently* devoted to active learning and that 36% of class time *should be* devoted to active learning (Fig. 5). These findings suggest that both groups of students want more active learning in their classrooms than currently experienced and desire a similar amount (~36% of class time) dedicated to active learning. Overall, most students have positive perceptions of active learning and perceive the benefits and/or utility of these practices. However, both student populations still valued listening to lecture (Fig. 7).

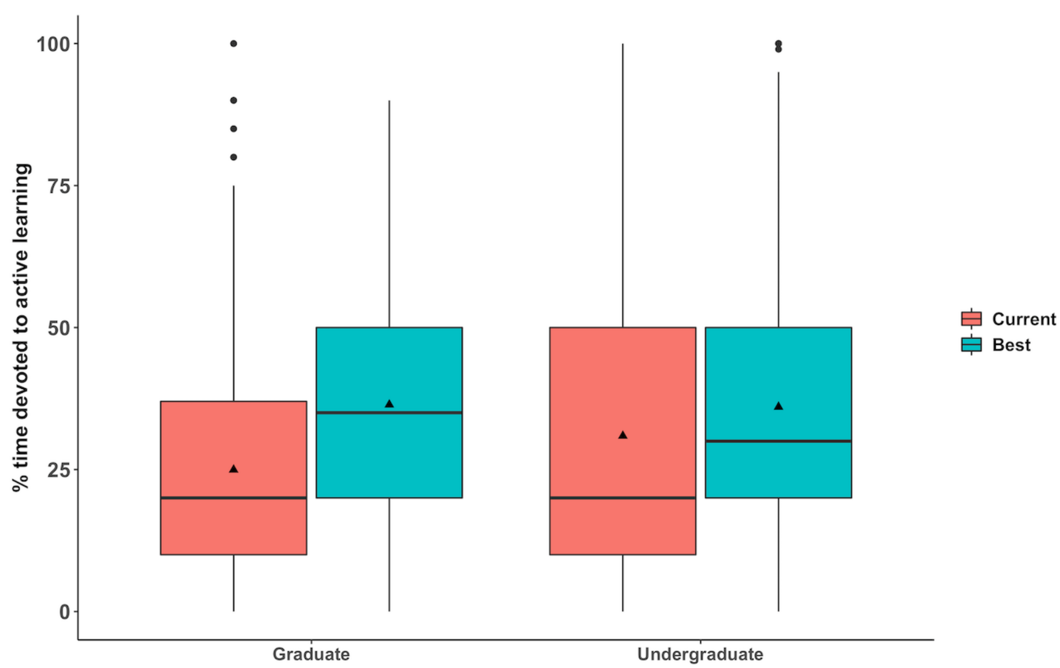


Figure 5. Box and whisker plots of the percent of class time graduate and undergraduate students think is currently ("Current") and should be ("Best") devoted to active learning. Students were provided a definition of active learning (Miller and Metz, 2014) and via an open-ended response, asked to estimate the amount of class time typically devoted to active learning and how much time they think *should* be devoted to active learning. Triangles indicate mean values. *indicates p-values ≤ .05.

Specifically, undergraduate students preferred instructional practices that involve peer-assisted learning and direct feedback from instructors (Fig. 6A). For undergraduate students, the top five most desired teaching practices were watching the instructor demonstrate how to solve problems (Q21), getting homework help directly from the instructor (Q10), brainstorming different solutions (Q2), studying course content with classmates outside of class (Q7), and asking the instructor questions during class (Q19). These preferences reflect that students value a variety of teaching strategies in their classrooms. For instance, students' understanding of conceptual questions increases after discussion with classmates regardless of students' initial knowledge of the answer (Smith *et al*, 2009). Undergraduate students also valued peer-assisted learning *outside* of the classroom and discussing course concepts with peers. These meaningful peer interactions outside of the classroom lead to gains in students' cognitive development (Jones *et al*, 2008) Similarly, graduate students' most desired forms of instruction included watching the instructor demonstrate how to solve problems (Q21), brainstorming solutions (Q2), asking the instructor questions during class (Q19), discussing concepts with classmates (Q9), and getting help from the instructor with their homework (Q10; Fig. 6A).

Figure 6. Box and whisker plots of the most and least desired instructional practices reported by graduate (*n*=161) and undergraduate (*n*=1113) students in STEM classrooms. (A) Most preferred activities. (B) Least preferred activities. Triangles indicate mean values. *indicates p-values $\leq .05$. Activities displayed are the top five most and least desired. Reported additional activities reflect the ranking misalignment between graduate and undergraduate students.

For undergraduate students, the five least desired forms of instruction were finding additional information not provided by the instructor (Q3), being graded based on the performance of a group (Q11), assuming individual responsibility for the learning material (Q8), making individual presentations to the class (Q5), and being graded on class participation (Q6; Fig. 6B). The least desired instructional practices were similar for graduate students: being graded on group performance (Q11) or class participation (Q6), finding additional information not provided by the instructor to complete assignments (Q3), working in assigned groups (Q4), and listening to the instructor lecture during class (Q1; Fig. 6B). Other studies have also found that undergraduate students felt unprepared to evaluate the value and importance of information and the work of others (Owens *et al*, 2017) and were often resistant to working collaboratively when their grades were on the line (Machemer and Crawford, 2007; Patrick *et al*, 2016; Owens *et al*, 2017). It is noteworthy that graduate students also disliked these teaching practices because in many ways these are vital elements of modern scientific practice. Using transparent grading rubrics, making expectations clear, and using best practices when assigning group work may help to increase student buy-in. The results for all 23 teaching practices can be found in the upper panel of Fig. S1. Although the mean values vary, the median desired level of each teaching practice mostly centered around *about the same*.

The summary statistics above highlight the trends in the data, but they also mask important variation that lends insights into student perceptions of active learning. Fig. 7 illustrates the variation in responses for three example teaching strategies (responses for the remaining teaching strategies can be found in S2-S21 Figs). As reflected in Fig. 5, most students experienced lecturing (Q1) the majority of the time. Very few students reported experiencing courses in which 70% or less of the time was devoted to lecture. Overall, students desired lecture about as much as they were currently getting (Fig. 7). This interest in lecture may seem counter to the documented benefits of active learning teaching practices (Freeman *et al*, 2014), however, a deeper look at the data gives the story more nuance.

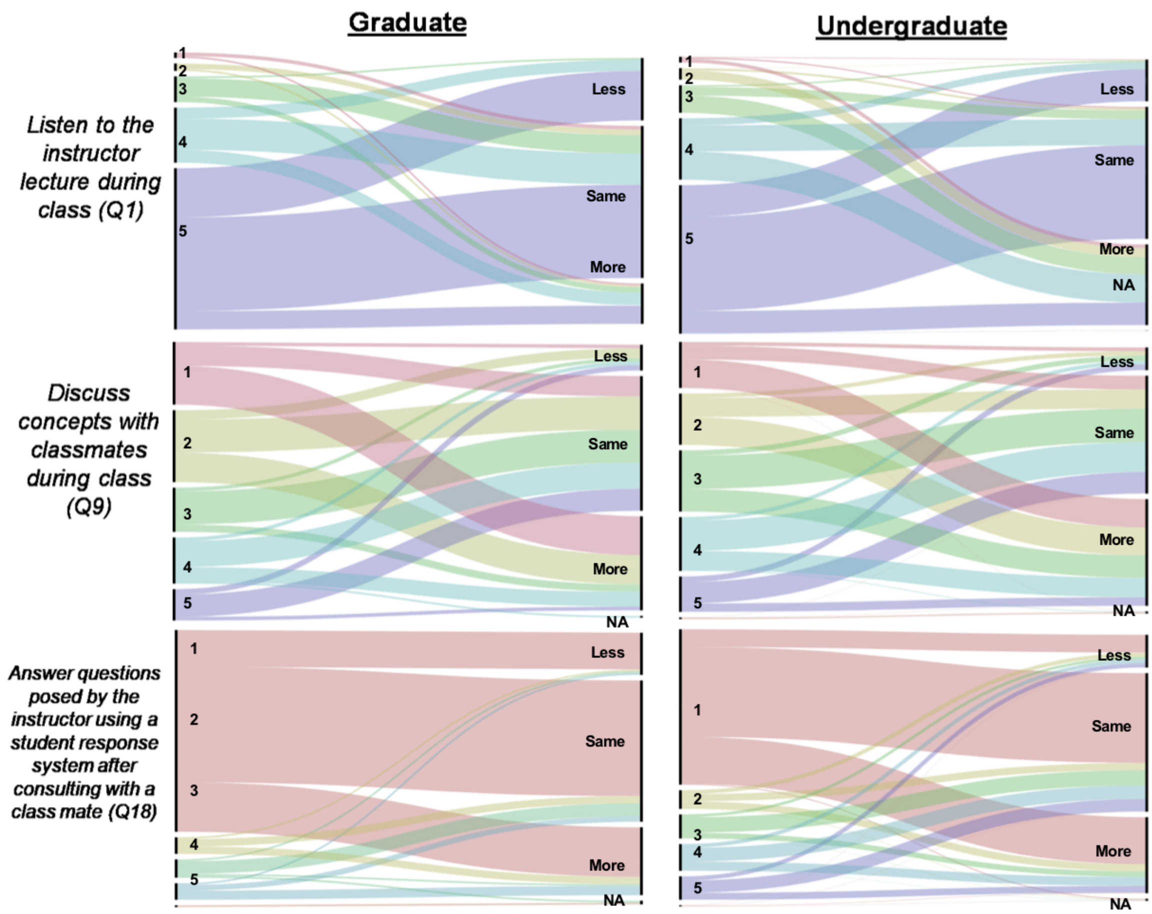


Figure 7. Alluvial plots of how often students reported experiencing a teaching practice and how much that teaching practice is desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left) and the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia," or lines between columns, indicate how often students reported experiencing a teaching practice (left side of each graph) and how much that teaching practice was desired (right side of each graph) and are proportional to the number of respondents. NA= respondents who did not respond to a particular question; 1=Never or almost never (0-10% of the time); 2=Seldom (11-30% of

the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time). Less=a combination of Much less and Slightly less desired; Same=About the same; More= a combination of Slightly more and Much more desired.

For example, approximately equal numbers of students reported that discussing concepts with classmates during class (Q9) happened *never* or *almost never* through *very often* (Fig. 7). Despite a median desired value of *same* (Fig. 6A), this particular active learning teaching practice was desired more often by a substantial number of students; very few students who *never or almost never* experienced this practice wanted less of it (Fig. 7). Answering questions posed by the instructor using a student response system after consulting with a classmate (Q18) was experienced *never* or *almost never* by the vast majority of the students in our sample (Fig. 7), and while the majority of students desired this teaching practice about the same amount, a sizeable proportion wanted more and fewer still wanted less (Fig. 7). Similar trends are evident for most of the remaining 20 teaching practices (S2-21 Figs).

The data above demonstrate that while many students value lecture, they also desire active learning practices, suggesting the students in our study “buy-in” to active learning—that is, they perceive it as valuable. However, we also found that no single teaching practice was universally desired – or not desired – suggesting to us that educators should employ a variety of active learning teaching practices in graduate and undergraduate classrooms alike.

Student Perceptions of Active Learning Based on Free-Response Items

Seven hundred sixty-eight students answered the prompt, “*Please describe your experiences with active learning in the classroom.*” We analyzed all the responses, which include 104 graduate and 664 undergraduate student responses. We identified these responses by the categories and subcategories described in the methods. Sample comments are included, along with the total number of responses in each category, in Table 4. Each student response was coded, line by line. As a result, some responses were a combination of values under the three categories. For example, 64% (422/664) of the undergraduate responses were in a single category, 29% (193/664) in two, and 7% (49/664) were in three categories. Likewise, 72% (75/104) of graduate responses were in a single category, 21% (22/104) in two, and ~7% (7/104) were in three categories. For both graduate and undergraduate students, most responses were in a single category. In Table 3, we provide an example of a response coded for more than a single category.

Table 3. Student response coding template/example.

Please describe your experiences with active learning in the classroom:	Categories	Subcategories
I currently have my biology class in an active learning classroom and one professor knows how to use the active learning classroom and it really benefits all the students, but the other doesn't know how to use it properly and I learn less from him than I do in a normal lecture without active learning.	**Positive:**	1. Active Learning is Beneficial
	Negative:	1. Professor does not implement active learning approach correctly

Table 4. Categories (Positive, Negative, and Constructive) and subcategories identified in student responses to the prompt "*Please describe your experiences with active learning in the classroom.*"

	Categories	Undergraduate Comments	Graduate Comments
Positive	Active learning makes the class engaging	"I personally prefer active learning in the classroom. It's difficult for me to focus during long lectures so having an activity to direct my attention to helps me better apply the material I have learned and understand what I need improvement on." (85)	"Active learning broke the class into small groups to work on a problem together. I felt I got to use the material rather than passively take it in…" (14)
	Active learning is beneficial	"In a previous math class, there was a significant amount of active learning, with very engaging exercises and significant in-class participation. Although I was off-put at first by how different the class was at first, I feel now as though I retained and	"Even graduate students benefit from active learning so I wish it was used more in these higher level courses!" (62)

		ultimately was much more successful in that course than in others due to the number of activities we went through.” (380)	
	Active learning helps with content retention	“When implemented correctly (not just clicker-question based), I find active learning to be helpful and good for long term information retention.” (88)	“Active learning classrooms are harder but I do retain and learn the information much better and still recall things years later.” (16)
	Active learning builds community	“[Active learning] can be a great way to connect with other students and build community.” (56)	“[Through active learning], I got to know my classmates better, which made coming to class much more enjoyable.” (7)
	Active learning helps with real world work environment	“I do not necessarily enjoy active learning and I often hate working in groups, but I understand that it is an important skill to learn and I understand why active learning is important.” (9)	“Active learning is essential as it helps to understand the concepts, retention of information and how to use the acquired knowledge in real life situation.” (1)
Negative	Active learning limits individual thinking and learning	“Active learning doesn't help me, personally, in the classroom. I learn best by watching, taking notes, and reviewing and quizzing myself on the material on my own time. I like things like clickers, to make sure I understand a concept, but otherwise, non-lecture classrooms keep me from focusing on what I need to learn.” (75)	“I am not a person that functions well in group settings. I learn much better by figuring out the problem on my own and discussing questions at my own pace with the lecturer.” (7)
	Professor does not teach using an active-learning approach correctly	“Sometimes it's helpful, sometimes it's not. It really depends on the engagement of the professor with the students and the discussions between group members.” (117)	“We are paying a lot of money for the expertise of the instructors and their presentation of the information through a traditional lecture. Active learning is guided in concept. But in practice it is mostly the students teaching themselves. And if I wanted that, to teach myself, or learn from other clueless students, I either wouldn't pay this much tuition or I wouldn't go to class.” (10)
	Active learning should only be used in certain fields	“In the introductory courses, I found active learning to be a great way to learn concepts through practice. However, while I have enjoyed its usage as a teaching method, I feel that active learning may not be the best teaching/learning strategy for every course” (25)	“I feel as though active learning is appropriate for some classes, but not in others. Sometimes I understand lecture material better when it is simply presented by the teacher, and other times active learning is necessary to better understand concepts. I think it really depends on the course.” (5)
	Current active-learning methods are a "waste of time"	“Active learning is so terrible. [Name of College] thinks it is the best thing in the universe, but there really is no benefit besides me hating my group. also, that basically means the professors don't actually have to teach, which means no learning is done besides busy work.” (98)	“Once you get to high-level (graduate-level) courses, these [active learning techniques] become a waste of time. You no longer need to "trick" highly motivated PhD students into learning.” (14)
Constructive	Active learning works when people are prepared to collaborate	“if a professor is not prepared well, it does not go well. If other students are not participating, it does not go well.” (50)	“In general, however, the best types of active learning for me are those that require me to work with the material independently (problem sets, games, written reflections, etc), and not necessarily debates or discussions. However this is mostly because I can't rely on my peers to have done enough homework or reading to maintain a valuable debate.” (16)
	Active learning only works in smaller classes.	“I have enjoyed the use of active learning in all my science classes especially the smaller classes. In my largest class it did feel like more of a way to track attendance and get points than to actually learn the material.” (16)	“I do NOT like activities for the sole purpose of having 'active learning...The active learning I did enjoy occured in small classes, when either questions were encouraged during lectures or we had to solve cases as a class with the professor as a resource.” (3)
	Effective active learning is	“I don't like how active learning is done currently in class, but I think the idea has potential. If it didn't involve so much busy work or multitasking I would	“The combination of very well-presented thorough material and explanations in lectures with a couple minutes (i.e. short activities) to work with the

	desirable	love it a lot more. Additionally, I don't think active learning should replace lectures, but instead work along side them." (116)	concepts/material independently was much, much, much more useful and helped me to learn." (18)
	Active learning works well when supplementing other methods	"When students have to go to several lecture based classes during the day, it is easy for them to start tuning out the teacher. If active learning were used in addition to the lecture, it would be harder for students to tune out because they have to be listening in preparation for an activity being done later in the class. Class time would be more enjoyable and engaging if some element of active learning were involved." (75)	"I think active learning can be useful, however, I think the combination of traditional lectures and labs accomplishes the needs of active learning. Often, active learning activities in the classroom don't result in the creation of useful documentation that can be consulted before exams" (6)

Note. Number of responses in each subcategory are included in parentheses after a sample comment.

The three overarching categories are Positive (in which the student is voicing an appreciation for active learning); Negative (in which the student is unambiguously negative); and Constructive (in which the student suggests the conditions under which active learning can be useful). We identified overwhelmingly positive codes for both graduate and undergraduate student responses, followed by negative and constructive codes (Table 5).

Table 5. Code sum total of identified categories.

Group Type	Positive	Negative	Constructive	Sum
Undergraduate	618	315	257	1190
Graduate	100	36	34	170

Five Positive subcategories emerged—active learning a) makes the class engaging, b) is beneficial, c) helps with content retention, d) builds community, and e) prepares for a real-world work environment (Table 4). Over half of both the graduate (62/104) and the undergraduate (380/669) students expressed a positive impression of active learning (subcategory "active learning is beneficial") (Table 4). One graduate student reported, "*I felt I got to use the material rather than passively take it in, and I got to know my classmates better, which made coming to class much more enjoyable*." An undergraduate said, "*Active learning is a very effective method to get students to really understand their curriculum since questions are encouraged as well as predictions/guesses.*" Many of the responses spoke to a general appreciation for active learning. However, some responses included disclaimers such as, "*I think a traditional lecture has its place explaining basic concepts and helping students get comfortable with the material, but activities can help engage and apply the material,*" (undergraduate student) and "*Sometimes [active learning] is effective, and sometimes it is patronizing,*" (graduate student).

Four Negative sub-categories were identified—(a) active learning limits individual thinking and learning, b) professor does not implement active learning approaches correctly, c) active learning should only be used in certain fields, d) current active learning methods are a "waste of time" (Table 4). In these negative subcategories, the largest number of comments belonged to "current active-learning methods are a 'waste of time'"—for both graduate (n=14 responses) and undergraduate (n=98 responses) students (Table 4). According to one graduate student, "*Some [professors] are awesome at using active learning for its positive, intended uses while others are totally off-base and waste class time without any benefit to the students.*" Similarly, an undergraduate student opined, "*The material has to be dumbed down so any idiot can figure it out, and working in groups means you work at the slowest pace of anyone there. It also doesn't help that there's barely any time for the professor to talk about a new concept.*"

Lastly, four Constructive sub-categories emerged—(a) active learning works when people are prepared to collaborate, b) active learning only works in smaller classes, c) effective active learning is desirable, d) active learning works well when supplementing other methods (Table 4). The most comments in this domain belong to the sub-category, "*Effective active learning is desired.*", evident in 116 undergraduate and 18 graduate student responses (Table 4). Student suggestions varied, from "*Sometimes they [active learning] feel[s] like a hassle. Good active learning should be integrated smoothly, but with a clear goal in mind,*" (undergraduate), and, "*Usually there is some lecture portion, then the active learning with worksheets or games, etc. What is most vital is that we actually have the information to be able to do the activity before we do the activity. I found it happens a lot where we have no idea what we are doing or the instructors have given no examples so we sit and play with our thumbs because we haven't learned anything. Learning first. Activities next to reinforce concepts and maybe expand upon them,*" (undergraduate), to "*it's helpful when it is done correctly. sometimes it's more distracting than helpful,*" (graduate), and, "*Often spent at least half of lecture time going over a paper in groups. Mostly effective for learning the paper if students actually read it, but there was no system in place for ensuring accountability, leading to some group sessions suffering. Also, while a lot was learned during these times, what was discussed in class did not usually help for questions on the exams,*" (graduate).

Notably, apparent differences between the two populations were not evident. Specifically, the themes that were the most common for undergraduate students were the most common for graduate students as well (Table 4).

Are There Notable Differences Between Undergraduate and Graduate Students in the Amount or Type of Active Learning Experienced or Desired?

Undergraduate and graduate students reported similar patterns in the most common instructional activities (Fig. 4A). Some differences emerged in the least common teaching practices (indicated by * in Fig. 4B), the most and least desired teaching practices (* in Fig. 6), and how much time was currently devoted to active learning (* in Fig. 5). While all students rarely experienced individual presentations (Q5), such practices occurred significantly more often in graduate than in undergraduate courses. Undergraduate students also reported significantly fewer opportunities for verbally answering questions in class (Q15) compared to graduate courses. For graduate classes, the use of student response systems (i.e.,Top Hat, clickers), either independently (Q17) or following consultation with a classmate (think-pair-share; Q18), was significantly less common than in undergraduate classrooms (Fig. 4B). Overall, graduate students reported slightly, though significantly, less class time devoted to active learning practices than their undergraduate counterparts (Fig. 5).

Although both groups of students wanted more direct engagement with instructors, undergraduates had a greater desire for such activities than their graduate counterparts. (Figs. 3A, 4; S9, S17 Figs). Specifically, undergraduate students sought more opportunities to ask their instructors questions, and seek information directly from instructors for homework (S9, S17 Figs). Compared to graduate students, undergraduates also wanted more opportunities to study with classmates outside of class time (Fig. 6A, S7 Fig). Furthermore, there was a significant difference between undergraduate and graduate students in the desire for less lecture (Q1). Graduate students identified lecturing as one of the top five least desired activities in their classrooms and wanted significantly less than undergraduate students (Figs. 3, 4). Finally, undergraduate students had a significantly less desire for assuming responsibility for learning material on their own, listing it as one of their top five least desired classroom activities (Fig. 6B, S8 Fig).

Overall, graduate student views are similar to those of undergraduates. Both groups of students want more active learning than they are currently getting. These results are similar to recent studies examining perceptions of active learning at a different university (Patrick *et al*, 2016, 2018), suggesting that these findings indicate a larger trend in higher education. Additionally, graduate and undergraduate courses implemented similar teaching strategies, dominated by lecturing; in fact, significantly less class time was devoted to active learning in graduate courses compared to undergraduate courses.

Conclusions and Implications

Any conclusions from these findings should be tempered with the limitations of this work. We hesitate to assign differences to our two populations beyond their status as either graduate or undergraduate students. Nonetheless, it is a reasonable assumption that graduate students are characterized by different levels of, for example, intrinsic motivation than their undergraduate counterparts. This differential motivation may lead to different expectations of or demands from their instructors. Future work, in which we attempt to align student responses with different aspects of student affect (e.g., motivation, mindset, self-efficacy) with perceptions of teaching strategies, would provide additional clarity. Further, we realize that graduate courses are likely to have a different culture than undergraduate courses, informed by factors beyond e.g., course level and class size. For example, instructors may perceive graduate students as closer to being colleagues than undergraduate students; these differences likely change in-class behaviors—of both students and instructors. It would be helpful, in follow-up work, to combine student perceptions with those of their professors.

Regardless of these and other unidentified limitations, our work contributes to the relatively small body of literature exploring the use of evidence-based, active learning techniques in undergraduate and graduate-level STEM courses. Similarly, we contribute to understanding student buy-in to active learning in the curriculum. Critically, our approach is sufficiently fine-grained to isolate which evidence-based techniques are in place and which of these techniques are desired by STEM students.

We found that graduate and undergraduate students want to experience a higher degree of active learning in their STEM classrooms than they currently experience (Fig. 5). Additionally, our open-ended responses indicate an overwhelmingly positive experience when we sought student insights on their experience with active learning (Tables 4, 5). Though some differences were identified in the specific type of active learning preferred, both graduate and undergraduate populations wanted more direct feedback (i.e., formative assessment) and a chance to learn in small groups. Further, all students wanted to learn through student presentations and direct engagement with the instructor using student response systems.

On average, graduate students wanted to engage more in individual learning compared to undergraduates. Our study suggests that graduate students experienced low levels of active learning, significantly less than their undergraduate counterparts. Our findings collectively provide evidence for educators, especially those wary of student resistance to change, that students buy-in to active learning. These findings along with student input via an open-ended response suggest that most students would like to experience more active learning instructional practices in their STEM classrooms.

Our results confirm that evidence-based teaching remains relatively scarce in graduate courses. However, this part of the STEM curriculum remains insufficiently explored from the perspectives of who uses active learning, what pedagogies graduate students prefer, and whether student preferences are in line with the evidence for what works best in the classroom. Few studies are investigating active learning practices for graduate classrooms. The existing studies suggest that graduate students hold active learning perceptions similar to those of undergraduates, that is, neutral through positive (Patrick *et al*, 2016, 2018). For example, graduate students in a flipped classroom performed better on exams than in traditional lectures but disliked the extra time necessary to prepare for class meetings (Tune *et al*, 2013b). Graduate students also value courses that implement a variety of active learning practices, especially when familiarized with the activities (Lopez and Gross, 2008; Jones *et al*, 2010; Miller and Metz, 2014). Combined with previous studies, our findings indicate a need to integrate active learning throughout all levels of the curriculum. While it may be essential to use active learning in first-year STEM courses, time, and resources should also be allocated to innovative teaching practices in upper-division and graduate-level courses.

Finally, we find no evidence that students, on average, are resistant to the implementation of techniques such as student response systems, opportunities for hands-on group work, and opportunities for direct interaction with the instructor. Instead, we identify evidence that students would prefer *more* active learning in their courses. Fortunately, instructors can implement many of these preferred active learning techniques into their existing courses with relative ease. For example, there are several excellent sources designed for educators to develop a toolkit of in-class assessment techniques such as classroom polling, short written reflections, and think-pair-share activities (Angelo and Cross, 1993; Fink, 2013; Tanner, 2013a). These suggestions, combined with an awareness of student preferences, may help instructors, teetering on the brink of adoption, to leap into active learning.

IV. COOPERATIVE LEARNING IN A BIOLOGY LABORATORY COURSE: EXPLORING GROUPING STRATEGIES (Study 2A)

Introduction

Experiences centered around peer group learning have demonstrated gains for students' academic performance (Hulleman and Harackiewicz, 2009; Patrick, 2013) while building on practices prevalent in the discipline. Working in small peer groups in the classroom has been linked to increased student achievement and better student perceptions and attitudes in college classrooms (Johnson *et al,* 1991, 2014; Smith *et al*, 2011. For instance, a study by Springer *et al,* (1999) analyzed studies on small-group learning in undergraduate science, math, engineering, and technology classes and found that students working in groups had higher academic achievement, better attitudes toward learning, and increased persistence in classwork compared with students in more traditional classes that lacked group work.

Positive experience in peer groups can also form the basis of a student's academic and social support early in their undergraduate experience. In fact, these support groups can extend beyond the course and continue as students progress with their major. The lasting bonds have a vital role in a student's success and persistence through the discipline (Johnson and Johnson, 1991). On the other hand, a negative peer experience may hinder a student's academic performance and harm a student's sense of belonging. A lowered sense of belonging is particularly harmful to first-year students, including having a negative consequence on students' persistence within the major (Hausmann *et al,* 2007). Thus, the STEM/biology teaching laboratory, which is exclusively driven by group work, has emerged as a key arena for studies involving retention (Corwin et al, 2018; Rodrigo *et al,* 2018), performance (Matz *et al,* 2017; Sato *et al,* 2014), and student sense of belonging in STEM (Esparaza *et al,* 2020).

There is considerable debate in the literature on how such positive peer groups should be formed in the classroom setting. Further, most of the research in this area is concentrated in the K-12 space (Agrawal *et al,* 2014; Kulik and Kulik, 1982; Slavin, 1987; Slavin, 1990). The limited research in higher education suggests no uniform consensus on grouping strategies (Donovan *et al,* 2018).

One form of group work is when members are assigned and groups are longer lasting as students work together in cooperative learning groups; students work towards a shared goal and are assessed individually with their groups (Johnson and Johnson, 1998; Tanner *et al*, 2003). Assigned long-lasting groups are among the three main types of cooperative learning, along with formal and informal learning groups (Johnson and Johnson, 1998). According to Johnson and Johnson (1998), *"assigned base groups are long-term heterogeneous groups with stable membership whose primary purpose is for members to give each other the support, help, encouragement, and assistance each needs to progress academically,"* (Johnson and Johnson, 1998).

Cooperative learning primarily stems from socio-constructivism. Based on the work of theorists including Lev Vygotsky, socio-constructivism recognizes learning as a social phenomenon (Vygotsky, 1968). In a socio-constructivist framework, learning requires students to conduct meaningful activities with peers as they think about what they are doing (Wankat and Oreovicz, 1993). Unlike regular group work, the instructor provides guidance and support to the peer groups throughout the task, beginning with forming groups in a cooperative learning environment. In a successful cooperative group, students grow to see each other as supports to their own ability to learn in the course. As a result, instructional environments including cooperative learning groups promote student-centered approaches and facilitate learners' co-construction of knowledge.

While there is consensus on the benefits of cooperative learning, there is less consensus on how best to form the groups to maximize learning and improve students' group learning experience (Donovan *et al,* 2018; Nhan and Nhan 2019) in the context of a laboratory course in a small liberal arts institution. Secondly, the nature of heterogeneity within groups can be based on a variety of characteristics including academic ability and social identities. For instance, in an upper-division neurobiology class, Gaudet and colleagues (2010) used students' self-reported academic major as a basis to form diverse learning groups, allowing transparency of group formation and for students to self-identify with a major. McInergney and Fink (2003) instead formed heterogeneous groups in their microbial physiology course by utilizing students' previous academic experience in microbiology and chemistry. Finally, in examining the literature on group work in higher education, there is limited evidence for the consideration of course interest in group heterogeneity. Yet in the first year of a STEM course, students have a greater variation in course interests along with varying academic characteristics than students in subsequent years (National Research Council, 2012).

A standard method for forming groups is to allow students to self-select their groups (observation, Chapman *et al,* 2006; Donovan *et al*, 2018; Nhan and Nhan, 2019). Students choose their group members in self-selected groups and have the autonomy to change groups during a semester (Springer *et al*, 1999). Self-selected groups require little planning on the instructor's part and may yield benefits as students group with their friends (Strong and Anderson, 1990, Bacon *et al*, 1999). As a result, group members may already have comfort with each other, and therefore, less class time and responsibility of instructors is required for structuring effective group practices. However, there are some drawbacks, especially in large enrollment classes. For example, Freeman *et al,* 2017 found that students tended to work with peers of the same ethnicity, gender, and similar academic ability. Such groups can lack the diversity of skills and different group members' perspectives (Mello, 1993; Bacon *et al*, 1998). In another study, Feichtner and Davis (1984) found that students self-report their worst group experiences when groups were self-selected. In contrast, Bacon *et al*, 1999, found that more students report their best experiences in courses where they had the autonomy to form their groups. The gains were particularly noticeable after a first term, where students had a chance to know each other outside of a course group environment (Bacon *et al*, 1999). The sum of the evidence is mixed for self-selected groups, yet this approach remains the conventional practice (Bacon, 1990; Strong and Anderson, 1990; Donovan *et al*, 2018).

The conflicting results in group formation on maximizing learning and student perceptions and attitudes towards group work leads me to explore how students experience working in different groups in a first-year course (Study 2A) and explore any possible impact on their academic performance (Study 2B). Evaluating student perception of their experience is essential because their perceptions affect their behaviors, their reaction to the environment, and their beliefs about the environment (Liskin-Gasparro, 1998; Williams and Burden, 1999). Therefore, indirectly or directly, student perceptions of their experiences in lab groups could be critical to their sense of belonging in STEM. Secondly, understanding the student experience of the varying grouping strategies can inform instructors on the promise of this pedagogical tool, informing them how best to utilize their limited time and effort on grouping strategies.

In Study 2A, I addressed how first-year students experienced two different methods of group formation while engaged in extended group work in a biology laboratory course. I explored how students experience learning in groups via self-selected groups, common in the discipline, and also via heterogeneous groups based on prior knowledge and interest in the subject. I also considered gender where possible to ensure students have gender allies within their groups and to ensure they are not the sole representatives of their gender within a group (Handelsman *et al*, 2004; Rosser, 1998). I used a pre-semester assessment to gauge the participants' prior knowledge of the current course content and their retention from material covered in the preceding course (Fig. 8, Appendix A). The pre-assessment established the criteria for the assigned groups. Interest in the course was determined from a Likert scale response in the pre-semester survey. The assessment questions were created by two faculty members who have taught the current and the preceding courses for over five years. A senior biology professor with extensive discipline-based education research (DBER) experience also verified the assessment.

Research Questions

My ***Research Questions*** for 2A are:

1) How is learning within assigned and self-selected base groups perceived by undergraduate students?
2) How did the perspectives of students change (if at all) after working in self-selected and assigned groups for a semester's duration?

My ***Research Questions*** for Study 2B are:

1) What, if any, are the elements of cooperative learning prevalent in self-selected and assigned groups in an introductory biology laboratory course?

2) How, if at all, does academic performance vary between students in the self-selected and assigned groups?

I hypothesized that students in both group types would show an improved attitude to group work over the course of a semester, as measured regarding a variety of cooperation-related factors. However, I expected that students in the heterogeneous assigned groups would experience a richer cooperative learning experience, reflected in a greater improvement in perceptions and attitudes to cooperative group work. Moreover, I expected assigned groups to result in improved academic performance over the semester and experience a higher frequency of cooperative learning elements as measured by the Cooperative Learning Observation Protocol (CLOP) tool (Study 2B).

Methodology

Context

Six sections of an introductory biology lab at a liberal arts college in the midwest were enrolled in this study. The lab sections participating in the study were each taught by two women instructors. Both instructors have taught this course over multiple semesters before. The participants consisted of students enrolled in the six sections (Table 6). Through an open-ended response, students self-reported their gender. In the assigned base groups, 72% of the study participants identified themselves as women, and 28% as men. In the self-selected sections, 60% identified as women, 27.5% as men, and 12.5% skipped the question (Table 6).

Table 6. Study participant attributes.

		Group Type			
		Self-selected		Assigned	
		(n)	(%)	(n)	(%)
Gender	Women	24	60%	33	72%
	Men	11	27.50%	13	28%
	No Answer	5	12.50%	0	0%
First Generation?	No	35	88%	38	83%
	Yes	5	13%	8	17%
Year in School	First Year	30	75%	36	78%
	Second Year	8	20%	8	17%
	Third Year	1	3%	1	2%
	Fourth Year	0	0%	1	2%
	No Answer	1	3%	0	0%
Ethnicity	Asian	3	8%	3	7%
	Black	1	3%	2	4%
	Hispanic	2	5%	1	2%
	White	28	70%	39	85%
	White & Asian	2	5%	0	0%
	No Answer	4	10%	1	2%

Survey Design

The pre-semester survey was designed to gather information on students' predictions about the course, their current perceptions, and attitudes about working in cooperative learning groups. The post-semester survey was designed to collect data on student perceptions and attitudes about working in cooperative learning groups after the course's commencement (Kern *et al*, 2007). The survey's open-ended responses were qualitatively coded using an inductive approach, and the quantitative responses were statistically analyzed. The purpose of the surveys is to learn how students enrolled in an introductory biology course valued working in groups, their apprehensions/concerns with group work, and to gauge their overall experience of a full semester of group work. The data collection type was through a questionnaire and administered online, (Fink, 2000; Nesbary, 2000; Sue and Ritter, 2007).

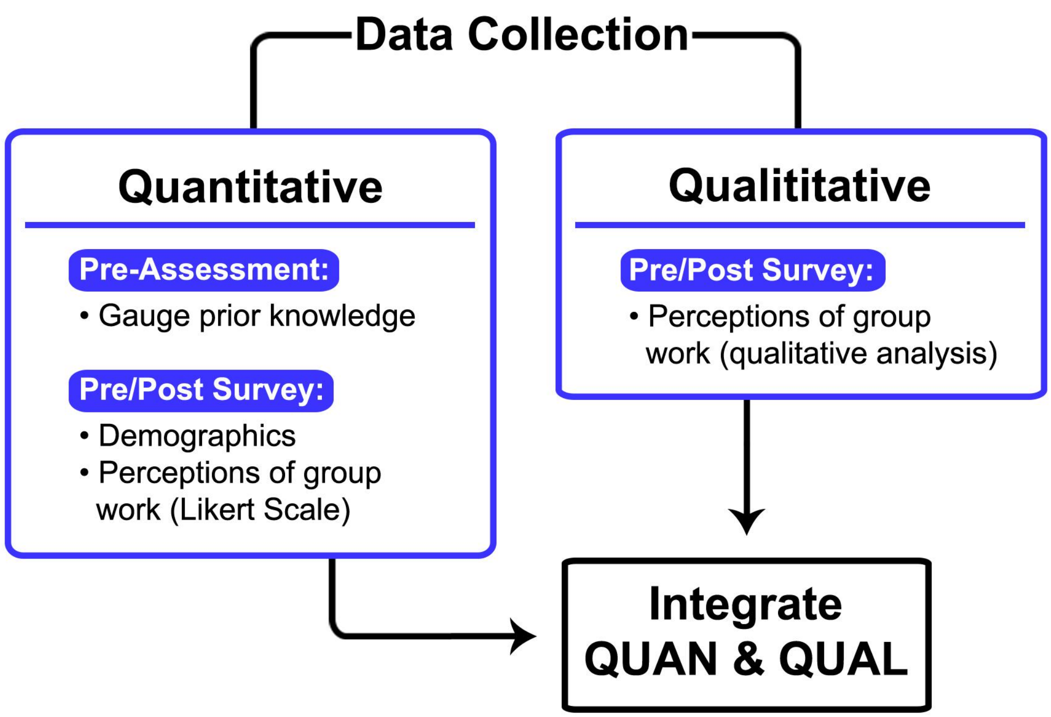


Figure 8: Overview of study data collection via a convergent mixed-methods approach.

Data Collection Tools

Consent: Instructors went over the consent form and answered any students' questions during the first lab section. During this time, students had the opportunity to ask questions they had regarding the study. The study was initiated once written consent was obtained from all participants and focused only on groups in which everybody had provided consent.

Surveys: Two surveys were conducted including a start-of-semester and end-of-term survey both validated by the think-aloud protocol (Fig. 8) (Charters, 2003). Student demographic data including their gender was determined through these questionnaires, along with their perceptions of the course and group work both at the beginning and end of the semester (Fig. 8). The start-of-semester survey included questions about their interest in the course. Questions related to interest were included since most of the participants are in their first year of college, yet to declare their major, and, therefore, likely have diverse interests towards a required course.

Google Forms was utilized to administer the survey. An officially constructed email was sent from me (Project PI) to all participants. The course instructor also reshared the email. The email detailed the overview and objectives of the survey. The survey details, including the study's purpose and goals, were restated in the first lab of the semester. To increase the response rate, a follow-up email was sent three days after the first lab's commencement.

Qualitative Data Analysis: Two cycles of inductive codings were analyzed by the coding team (2 coders) for all open-ended qualitative responses in the pre and post-semester surveys (Fig. 8) (Saldana, 2009). Each of the open-ended responses was read through by the two *in-vivo* trained coders both unaware of whether responses were from students in the assigned or self-selected groups (Saldana, 2009). Via team consensus, a subcategories codebook of all issues that recurred in the data was generated (Charmaz, 2006, p. 55). This approach allowed the subcategories to emerge from the data. Once the team generated a subcategories book, 10% of the comments were selected to determine interrater reliability via Cohen's Kappa (κ). This is an established statistical reliability test. Raters received κ =.94, a *strong* agreement on the Kappa scale (Cohen, 1960). Once reliability was established the remaining responses were equally divided among the two coders. Once all the data was coded into subcategories, patterns and connections within and between the subcategories were examined. Key ideas expressed within the subcategories were also identified, marking how participants respond similarly to each other and how their responses differ. Similar subcategories were sorted together into larger categories (Gonsar *et al,* 2021). In the results section, a summary of each category and subcategories that describe these points are included.

To identify which categories are most common across responses, the number of times particular subcategories came up were counted. These counts provide an estimate of relative importance, revealing the general patterns in the data. Since subcategories that were not within the recommendation for the survey were disregarded, all qualitative responses reported are within one standard deviation from the average (Baruch, 1999). Upon completion of all coding, all subcategories were tallied and the responses decoded to compare across the two group types.

Trustworthiness in Qualitative Analysis: Multiple approaches were employed to strengthen the rigor of the qualitative work (Lane, *et al*, 2019). Transparency was maximized for all the analysis and methods of qualitative coding to increase confidence in the analysis for future replication of this approach (Denzen, 1978). Each of the qualitative analyses was also completed as a team, arriving at all categories via group consensus, allowing all members to critically examine all the data and their interpretations (Eby *et al*, 2009).

Quantitative Data Analysis: A ten quadrant 2 X 5 contingency table for a two-tailed Fisher's Exact Test was created to determine differences in course and group work interest between self-selected and assigned groups, A Mann-Whitney U Test of paired differences, a nonparametric test for independent samples was run to compare cooperative related items between group types. A Wilcoxon Signed-Rank Test, a nonparametric test, was completed for all paired samples to compare within-group types.

All data remains stored in protected digital files and with an appropriate name system for each data type (Bazeley, 2013). The study design and survey instruments were approved by the institution's IRB (approval #1819-0153).

Results

Analysis

Below is a description of the quantitative statistical findings and the qualitative thematic results. Once the qualitative and quantitative analyses were completed, the data sets were compared to each other. There are areas where the data are in agreement and areas of disagreement on specific concepts in both sets of results. A combination of the qualitative and the quantitative sets of results provides the development of a more complete picture of assigned base groups and self-selected groups in introductory biology labs.

<u>Similar levels of course interest and preference for group work in early semester for both group types</u>

At the beginning of the semester, there was no statistical evidence for an association between group type and interest in group work (p=.649, Fisher's Exact Test). Around 68% of students in the self-selected group and 72% of students in the assigned group began the semester with a positive or somewhat positive view on group work with response rates of 80% and 93.5% respectively. Likewise, students in both groups had similar levels of interest in the course at the beginning of the semester (p=.764, Fisher's Exact Test), with 85% of students in the self-selected group and 95.7% in the assigned group responding to this question (Fig. 9).

Frequency of students working with each other

______At the end of the semester, students confirmed how often they worked with each other (Fig. 9). Students reported statistically similar results (p=.33, Mann-Whitney). On a scale of 1-3 (Sometimes to Almost/All the time), assigned students on average worked together 2.07 and self-selected students for 1.95 as per the scale.

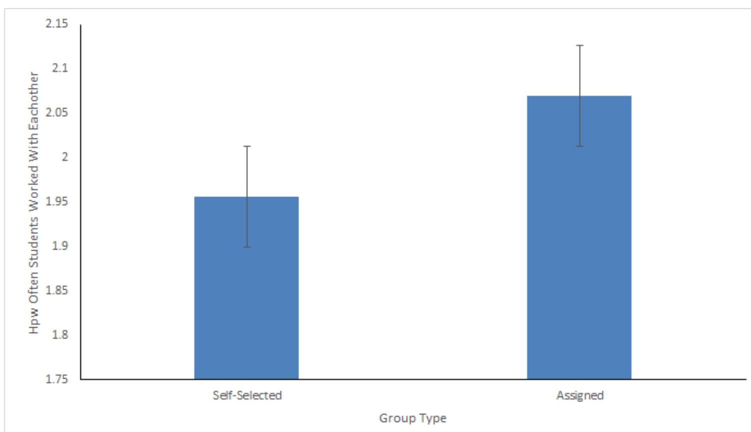


Figure 9: Frequency of how often students in the self-selected and assigned groups worked together. Students were asked to rate how often on a scale of 1-3 (Sometimes to Almost/All the time), they worked with each of their group members.

Collaboration and Understanding Material are Seen as Key Functions of Group Work

Students in assigned and self-selected groups responded to what they perceived as key functions of working in their group. In an analysis of pre and post-semester responses, codes were identified for the following categories: a) *understanding material,*

b) *collaboration skills,* c) *personal growth,* d) *negative/constructive experiences* and e) *bonding experiences*. For both group types, codes for *collaboration* and *understanding material* showed up most often both at the beginning and the end of the semester. (S Table 1). Responses regarding *understanding material* saw gains regardless of the group type. Students in both group types identified *collaboration*-related skills as the most frequently identified utility of group work at the beginning of the semester. For instance, when probed about the utility of their group, a student in the assigned group said, *"...is important in order to collaborate with others and gain a broader perspective of information."*

At the beginning of the semester, students in the self-selected group identified codes in the *collaboration skills* category to have the highest number of codes (27.5%), followed shortly by *understanding material* (25%), *personal growth* (17.5%), and finally by *negative/constructive experiences* (15%) and *bonding experience* (15%) (S Table 1). By the end of the semester, we found that almost 50% of the self-selected students identified codes related to understanding the material as the most identified utility of their group. This code was followed by *collaboration* and *personal growth*, each at 23.5%. There were minimal *negative* and *bonding experience* examples, both at less than 3% of all codes. Although self-selected students more frequently identified *collaboration skills* at the beginning of the semester, they found *understanding the material* to be a more frequent group work function by the end of the semester, followed closely by gains in *collaboration skills*.

Similarly, students in the assigned group identified *collaboration* as a utility of group work at the beginning and end of a semester. A student at the beginning of the semester said, *"I enjoy group work because different perspectives on a topic are easily accessible."* 34% of codes at the beginning of the semester related to *collaboration*. By the end of the semester, *collaboration*-related codes remained high on the ranking, but the percentage decreased to 26.5%. Students in the assigned groups continued to find utility in *understanding material*, which jumped from 20.5 % to 37.5% by the end of the semester. One student in the assigned group included, *"I have a great overall view on group work because my group members who sometimes have clarification questions help solidify my own understanding of what I do and do not understand."* Codes related to *personal growth* saw no change, and there were decreases in codes related to both *negative/constructive* codes. *Negative/constructive* related codes decreased from 18% to 9.4%. Like the self-selected groups, students in the assigned groups identified a similar ranking of categories at the beginning of the semester.

Pacing the Learning, Unequal Work, and Communication Issues are Key Challenges

Overall, there was approximately half the number of challenges identified compared to benefits for both types of groups (S Table 2). There were 51 benefits and 23 challenge codes for students in the assigned group. In comparison, there were 35 benefit codes and 20 challenges for students in the self-selected group. Fifteen respondents, or 40.5% of students in the assigned group and 33% (9 students) in the self-selected group, indicated no challenge in their group (Table 7). From the remaining responses, there were challenges across nine categories. These include: 1) *all members lack understanding of concepts* 2) *issues with communication* 3) *unequal work contribution* 4) *pacing learning and working* 5) *coordinating work outside class time* 6) *issues with group members* 7) *group gets sidetracked* 8) *members' lectures were at different paces* 9) *different working styles.*

The top 3 challenges in the assigned groups include a) *pacing learning and working while in a group*. The subsequent highest challenges were b) *unequal contributions* and c) *issues with communication* (Table 7). For instance, a student in the assigned group type said that the disadvantage of their group is that *"everyone is learning differently and working at different paces which can be frustrating."* The self-selected group identified similar challenges, however, in a different order of frequency. For instance, in the self-selected group, *unequal contribution* was the most identified challenge. A student reported, *"...I felt like I was always doing the work....."* Following this was the challenge of *issues with communication.* For some students in the self-selected group, the number of group members appeared to change from week to week. On the other hand, assigned group members had fixed 3-4 students in each base group. An assigned student expressed this challenge as, *"When the whole bench worked together, things ended up getting muddled and complicated."*

Table 7. Category (Challenges) and subcategories identified in student post-semester responses to the prompt "*From your perspective, what were the challenges associated with your group*?"

	Subcategories	Assigned Groups	Self-Selected Groups
CHALLENGES	Collective lack of understanding of concepts:	None	"sometimes we were all lost on a topic together, no one knew how to move past it." (3)
	Issues with communication:	"...challenges included coming to a consensus." (4)	"...When the whole bench worked together, things ended up getting muddled and complicated." (4)
	Unequal contribution:	"...work not always equal" (4)	"...I felt like I was always doing the work and felt bad if I tried being in another group," (6)
	Pacing learning and working in a group:	"...disadvantages are that everyone learns differently and works at different paces which can be frustrating" (6)	"some challenges were the speeds we wanted or could work at were somewhat different which made staying together and helping one another hard at times" (3)
	Coordinating work outside class time:	"...challenges were finding time to meet outside of class." (2)	"but it's always hard organizing time outside of class to finishe group projects." (1)
	Issues with group members:	"some were hard to work with" (1)	"Some challenges were that the other two in my group were close friends and I was not so it was sometimes hard to break into their discussion." (2)
	Group gets sidetracked:	"The only challenges was when we got towards the end of the year when we were more comfortable with each other, we would get distracted easily as we enjoyed each others company. However, we were able to get back to work quickly as well." (2)	"I sometimes worked with groups of people I was already friends with and it was difficult to stay focused." (1)
	Group members' lectures were at a different pace:	"The only challenge we had was the different paces we were all at in our classes, as some of the lab was covering material two of us group members had not yet reached in class." (2)	None
	Different working styles:	"...slightly different working styles, which occasionally clashed." (2)	None
	No challenges:	15	9

Note. The number of responses in each subcategory is included in parentheses after a sample comment.

Concerns Related to Group Work Diminish as Students Work Together

Students in both group types responded to what they perceived as concerns of working in groups at the beginning and again at the end of the semester. The categories identified from the responses include a) *unequal work/participation* b) *mindset,* c) *interest incongruence,* d) *effect on course grades* e) *hinder learning* and f) *group dynamics* (S Table 3). For students in the assigned groups, the top concern was *unequal work/participation* followed by *group dynamics.* Regarding *group dynamics, an* assigned student stated*,* "*Only concern I have is being paired with people who don't work well in a group because then it's counterproductive. Some people are reluctant to help others or share information and instead judgemental, which can make lab a negative experience. I don't expect to have that experience at [current institution] but in high school it occurred more than once.*" Self-selected students also identified *Unequal work/participation* as a key concern at the beginning of the semester. The second most frequently coded category of concern arising for self-selected students was *interest incongruence* (18.18%). This category included codes that captured concerns about group members with lower interest, who are perceived by peers to not care about learning or are perceived to be unmotivated. A self-selected student noted their concern, "*When there are people who don't care and don't want to help.*" In the assigned group, all but four students expressed some concern with working in their group. In the self-select group, all students expressed at least one concern about working in their group.

Unequal work: Similar to the key *challenges* identified, selected and assigned groups early in the semester perceived *unequal work/participation* as the primary *concern* about working in a group. 60.63% of codes from self-selected students and

40.63% from students in the assigned group expressed this concern (S Table 3). One student in the pre-survey stated, *"There's usually a group member who gives less effort and others have to make up for it."* Other concerns were the *effect on grades* and general *group dynamics* for students in the assigned group, each with 15.64% of the initial codes.

Following initial concerns regarding *unequal work/participation*, at the semester's end both groups of students indicated a statistically similar *equal contribution to the completion of the activity* in their Likert scale responses (p=.900, Fisher's Exact Test). In the post semester survey, assigned groups had an average response of 4.68, and self-selected students had a mean value of 4.56 (near the Likert scale value of *5=strongly agree*). 100% of all students in the assigned group and 97% of all responses in the self-selected group indicate they made an equal contribution to the completion of the group activity. I also found that students overwhelmingly agree that all group members made an equal contribution to completing the activity (p=.282, Fisher's Exact Test). I found a 93% proportion of agreement in the assigned group compared to 86% of self-selected students.

When students were again queried in open-ended responses about their key concerns at the end of the semester, nearly 50% of the assigned group codes were associated with *no concerns*, up from 12.50%. Likewise, I found 37.50% *no concern* codes in the self-selected students. Following a similar trend, codes related to *unequal participation* diminished in the assigned group from 40.65% in the pre-semester survey to 16.00% of the codes in the post-semester survey. In the self-selected students, codes in the *unequal participation* category decreased to 25.00% from 63.64%. To illustrate, a student in the assigned group stated the following at the beginning of the semester, *"If not* *everyone is willing to do their part, it affects the whole group."* By the end of the semester, when asked regarding their concern, the same student stated, *"I do not have any concerns."*

Assignments based on group performance: Students were also queried via a Likert scale response (*1=strong dislike, 5= strongly like*) regarding their feelings about being graded as a group at the beginning and again at the end of the semester. Self-selected students (p=.001 Wilcoxon Signed-Rank Test) and assigned students (p<.001 Wilcoxon Signed-Rank Test) both identified increased comfort with being graded as a group after the duration of the semester (Table 8). Responses for the assigned group shifted from 3.11 to 3.95 over the course of the semester, and in the self-selected group students, I saw a change from 3.10 to 3.70 (Table 9). I also compared the two group types to each other relative to their gains in preference for assignments based on group work. I found no significant difference (p=.592, Mann-Whitney U Test) when comparing the two group types (Table 9).

After their responses to the quantitative scale, students in the pre-semester survey expanded upon their responses to assignments based on group performance. Like the quantitative data, a close analysis of the open-ended coding also reveals similarities in both groups. Codes from the responses were bucketed in positive, negative, and constructive categories (S Table 4) (Gonsar *et al*, 2021). I found that for both group types, negative codes were dominant, followed by constructive codes and finally positive codes. I found 20 negative codes in the assigned groups and 12 for the self-selected groups. In comparison, I found 9 and 8 positive codes for self-selected and assigned groups, respectively. For both group types, most of the negative codes are overwhelmingly in the subcategory, *uncomfortable with grades based on other's performance*. For instance, a student in their assigned group said, *"I don't think my grade/effort should be impacted by the effort of another student because we could have very different workloads/work ethics."*

The shift in attitude for group grades was one of their top three gains for all the cooperative-related items measured; the most significant gain for students in the assigned group and third-ranked for self-selected students (Tables 8, 9; S22 Fig). Overall, students increase their comfort with group grading regardless of group type. Specifically, the pattern in the data suggests there is increased comfort across the duration of the semester with having assignments graded as a group, with a larger, although not statistically significant difference, in students placed in the assigned groups.

Table 8. Wilcoxon Signed-Rank Test of differences from pre to post semester survey item scales.

Activity	Self-Selected			Assigned		
	Pre	Post	**p-value**	Pre	Post	**p-value**
Enjoyment in Working in a Group	3.70	4.40	.001	3.73	4.46	<.001
Ability to Work Productively as a Group	4.34	4.28	.364	4.32	4.43	.173
Accountability for Effort/Understanding	4.44	4.72	.034	4.76	4.68	.224
Comfort Arguing Perspectives	3.66	4.16	.002	3.70	4.38	.001
Group Members Help to Develop New Perspectives	4.34	4.47	.154	4.38	4.59	.038
Group Work Enhances Understanding	4.22	4.56	.017	4.19	4.54	.015
Feelings Regarding Group Grades	3.10	3.77	.001	3.11	3.95	<.001
Comfort Asking for Help	4.07	4.60	.001	4.08	4.59	.002

Note. Students were asked to rate the following activities on a scale of 1-5 (*Strongly Agree-Strongly Disagree* or *Strong Like or Strongly Dislike*) at the beginning of the semester (Pre) and then again at the end of the semester (Post).

Table 9. Mann-Whitney U Test of individual paired differences (pre and post semester) between Assigned and Self-Selected students.

Activity	Self-Selected		Assigned		Paired Difference		p-value
	Pre	Post	Pre	Post	Self-Selected	Assigned	
Enjoyment in Working in a Group	3.70	4.40	3.73	4.46	0.72	0.77	.952
Ability to Work Productively as a Group	4.34	4.28	4.32	4.43	-0.06	0.11	.294
Accountability for Effort/Understanding	4.44	4.72	4.76	4.68	0.28	-0.08	.091
Comfort Arguing Perspectives	3.66	4.16	3.70	4.38	0.50	0.68	.388
Group Members Help to Develop New Perspectives	4.34	4.47	4.38	4.59	0.13	0.22	.779
Group Work Enhances Understanding	4.22	4.56	4.19	4.54	0.34	0.35	.849
Feelings Regarding Group Grades	3.10	3.77	3.11	3.95	0.69	0.84	.542
Comfort Asking for Help	4.07	4.60	4.08	4.59	0.50	0.51	.952

Note. Students were asked to rate the following activities on a scale of 1-5 (*Strongly Agree-Strongly Disagree* or *Strong Like or Strongly Dislike*) at the beginning of the semester (Pre) and then again at the end of the semester (Post).

Students Identify Social and Academic Skills as Key Benefits of Group Work

Students in both group types responded to what they perceived as the benefit(s) and the challenge(s) of their group type in an end-of-semester survey (Table 10, S Table 2). Twenty-seven students from the self-selected groups and 37 from the assigned groups responded to the prompt. Eight Beneficial subcategories emerged, including 1) *cohesive understanding of the course concepts,* 2) *contributing different ideas,* 3) *sense of accountability,* 4) *positive feedback between group members* 5) *enhanced own learning experience,* 6) *enjoyable learning environment,* 7) *developed new relationships,* 8) *all* *members contributed equally*. Amongst both group types, there were a total of 86 codes associated with benefits, including 51 codes for the assigned group and 35 for the self-selected group (S Table 2).

The top Beneficial subcategory seen from working in assigned groups was *contributed to generating different ideas* (17). Students in the assigned group also highly valued *positive feedback between members* as a key benefit to their group type, the second most coded subcategory. Exemplifying this, one assigned student noted, *"The benefits of working in a group throughout the semester were that we were able to help each other out and answer each other's questions when any came up."* Another student noted, *"...Great benefits from collaborative problem solving and additional feedback and insights."* Yet another student in an assigned group wrote, *"I was able to have questions answered from different perspectives and practice answering questions as well."* Rounding out the top third subcategory for Benefits among assigned students was that working in their groups was an *enjoyable learning environment* (9). For instance one student noted, *"I really enjoyed our group. we got along really well."* Another assigned student wrote, *"It was great to not have to choose people to be with all the time and I think our group got alone (along) really well."*

Likewise, in the self-selected group, the top 3 subcategories included group members *contributing to different ideas* (15), group learning being an *enjoyable experience* (5), and *positive feedback* between all members (5). Regarding *contributing different ideas*, a self-selected student noted, *"Having multiple people to bounce ideas off of was very beneficial but most groups were formed based on the people sitting near you."* A self-selected student who found *enjoyment* as the primary benefit of their group stated, *"It was pretty great to be able to work with people I already knew and could work well with."* Finally, encompassing the code for *positive feedback* from peers, a self-selected student wrote, *"recieve help from peers when not understanding."*

There were similar patterns in the benefits for students in both assigned and self-selected groups at the end of the semester. All students in the assigned and the self-selected groups reported at least one benefit from their group type. Overall, students found considerably more benefits than challenges to working in their groups regardless of group type (S Table 2). More than 50% of students in the self-selected group and 46% in the assigned group identified *contributing to different ideas* as a primary benefit of working in their groups (Table 10) the most frequently cited benefit of working in their respective groups. Both groups also found it essential that working with their peers made their laboratory experience more *enjoyable*.

Table 10. Category (Benefits) and subcategories identified in student post semester responses to the prompt, "*From your perspective, what were the benefits associated with your group*?"

	Subcategories	Assigned Groups	Self-Selected Groups
BENEFITS	Cohesive understanding of the course concepts:	"...had a better understanding of the material." "...deeper understanding of content and being able to discuss material with one another." (4)	None
	Contributing different ideas:	"...bounce ideas off of peers and to work through challenging topics with other classmates" "...to explore different ideas with your group and collaborate to find a reasonable answer." (17)	"Having multiple people to bounce ideas off of was very beneficial but most groups were formed based on the people sitting near you. " "The benefits of working as a group was the ability for each of us to bounce ideas of each other." (15)

	Sense of accountability:	"benefits included a strong flow of ideas, a sense of accountability, and a collective sense of togetherness" "...group had many benefits, which included sharing and collaborating ideas, especially since we all have different lecture professors, and encouraging and learning in lab." (3)	None
	Positive feed-back between group members:	"I was able to have questions answered from different perspectives and practice answering questions as well." "The benefits of working in a group throughout the semester were that we were able to help each other out and answer each other's questions when any came up." "Great benefits from collaborative problem solving and additional feedback and insights..." (9)	"you are able to work together if a problem is challenging, and we were able to work together to get tasks completed." "recieve help from from peers when not understanding." (5)
	Enhanced own learning experience:	"...Overall, having a group helped me (personally) to gain a better understanding of the course material." (5)	"was always willing to help me if I did not understand something." (5)
	Enjoyable learning environment:	"I really enjoyed our group. we got along really well.." "It was great to not have to choose people to be with all the time and I think our group got alone (along) really well..." (9)	"It was pretty great to be able to work with people I already knew and could work well with" "...a benefit is that you are able to work with those that you are comfortable with." (6)
	Developed new relationships:	"gained new relationships..." (2)	"I enjoyed working with my lab partner, we have become friends where we didn't know each other before this class." (1)
	All members contributed equally:	"When one of us didn't know the answer to a question or what to do, another would step in and help; there were very few times when we were at a total loss of what to do. No one argued or got into disagreements, and we were able to work efficiently because we divided work amongst ourselves and talked out issues." (2)	"...The amount of work we did was usually divided evenly." (3)
No benefits:		None	None

Note. Number of responses in each subcategory are included in parentheses after a sample comment.

Students Gain Comfort in Arguing Perspectives and Seeking Help over the Course of a Semester

I also evaluated individual changes for each student from pre to post-semester including their *comfort in arguing perspectives* via a Likert scale response, which showed an increase for students in both group types (Tables 8, 9, S22 Fig). The average quantitative scale for self-selected students was 3.66 at the start of the semester, with a significant increase (p=.002, Wilcoxon Signed-Rank Test) to a mean value of 4.16 by the end of the semester (Table 8). Likewise, the mean value increased from 3.70 to 4.38 for students in the assigned group. The increase was also statistically significant (p=.001, Wilcoxon Signed-Rank Test), making it one of the top 3 gains for the assigned group (Table 8, S22 Fig). When the gains were compared between the two group types, I observed no statistically significant difference (p=.388, Mann-Whitney U Test).

Working in both group types also increased students' *comfort in asking for help*. Students in the assigned group began with a mean value of 4.08 at the start of the semester. The mean value increased to 4.59 by the semester's end, which was significant (p=.002, Wilcoxon Signed-Rank Test). I found a similar pattern for students in the self-selected groups with a statistically significant mean increase from 4.07 to 4.60 (p=.001, Wilcoxon Signed-Rank Test) (Table 8, S22 Fig). Again I found statistically similar gains by the end of the semester when both groups were compared to each other (p=.952, Mann-Whitney U Test).

Students Saw Group Work as Increasing Their Understanding of Course Content

From the Likert scale quantitative responses, regardless of group type, students indicated an increase in content understanding over the semester. Self-selected groups (p=.017, Wilcoxon Signed-Rank Test) increased from 4.22 to 4.56 on average. Likewise, students in the assigned group (p=.015, Wilcoxon Signed-Rank Test) increased from a 4.19 average to 4.54. Both group types identified a significant increase from pre to post-semester (Table 8), and again there was no statistically significant difference between the group types (p=.849 Mann-Whitney U Test) (Table 8).

Similar patterns are also evident in the qualitative responses for both group types (S Table 5A-C). In the start-of-semester survey, 19 of the 22 self-selected responses agreed that working in their group increased their understanding of the material (S Table 5A). For instance, a self-selected student wrote, "*Having people explain how they think of problems and solutions helps me better understand questions. It also gives me a larger tool box to work out of,"* (S Table 5C). Assigned group students indicated a similar response when asked about whether working in their group increased their understanding of the material. Out of 34 responses, 26 agreed, five disagreed, and three neither agreed nor disagreed (S Table 5A). When asked to elaborate on their quantitative item scale, a typical response in the pre-semester survey included, *"My partner(s)* (previous course) *helped clarify things I was confused about, and made it more motivating to work hard because they were also,"* (S Table 5C).

Both groups overwhelmingly agreed that their group enhanced their content understanding by the end of the semester, mirroring the quantitative data (S Table 5B). Sixteen out of 18 responses from the self-selected group agreed that their group enhanced their understanding of the laboratory material. For instance, when the previously mentioned self-selected student was asked the same question in the post-semester survey, she wrote, *"There were many times when I had a group member explain a different point* *of view to me where the point of view helped me understand the material in a different way,"* (S Table 5C). For the assigned group, 24 out of 27 agreed that their group increased their course understanding. In the post-semester survey, the same assigned student quoted above said, *"Yes, I understood concepts a lot better with group work because if I didn't understand one perspective, another person would try to explain it to me,"* (S Table 5C).

Discussion

Consistent with my hypothesis, assigned and self-selected students identified improved perceptions and attitudes with group work, as measured from their open-ended responses and through cooperative learning factors. My findings suggest that self-selected and assigned group students overwhelmingly found more positive functions and benefits by working in their group type. Substantial benefits were *contributing different perspectives*, better *understanding the material*, and that working in their groups made labs a more *enjoyable experience*. Both types of groups also identified *learning to collaborate* and *gaining knowledge* as critical functions of their group. Overall, students perceived a few challenges or concerns from working in groups. I found that regardless of group type, the concerns related to group work diminish as students continue to work together. Specifically, regarding *unequal participation*, I found that not only did it decrease as a directly rated personal concern, but students also overwhelmingly rated both themselves and each other as contributing equally by the end of the semester.

Inconsistent with my hypothesis, both group types found similar functions and benefits in the cooperative learning experience, reflected in similar improvement in their perceptions to cooperative group work. My quantitative data revealed no meaningful differences between the two groups types, particularly in cooperative related items self-rated on a Likert scale. Both students identified gains in multiple cooperative social and academic-related skills during the semester within each group type. When I compared the two group types to each other, I found that the gains were statistically similar in the quantitative data (Table 9). When allowed to expand on their ratings using student free response, the coding also revealed similar categories dominating between the two groups.

The qualitative and the quantitative data address similar positive gains from group work for both assigned and self-selected groups. One possible explanation for the apparent lack of differences is that, on average, self-selected groups largely remained in the same groups for the semester's duration (Fig. 9). As a result of this stability, the self-selected groups, similar to the assigned groups, may have developed similar levels of cooperation. Studies by Johnson and Johnson suggest group stability as a factor that allows students to learn skills they need to resolve problems in working with each other (Johnson *et al*, 1994), and students were probed on how often they work with each other, students in both group types reported a statistically similar amount of time (Fig. 9). As such it is possible that the benefits of working in small groups may have arisen organically for self-selected students by working within stable groups for extended periods of time. For instance, there were high levels of cooperation and student satisfaction with the self-selection of groups in a study of 3rd and 4th-year university students who knew each other well (Nhan and Nhan, 2019).

However, though not evident quantitatively, a close examination of the qualitative responses reveals that students in the assigned group might have a richer positive group experience than students in the self-selected group. Specifically, I observed a higher proportion of codes associated with benefits and a lack of challenges for assigned groups (S Table 2). It is also notable that in 7 out of the 8 items queried the assigned groups had larger marginal gains, though not statistically significant (Table 8). This may suggest a substantial congruence insinuating that there may be a small effect from the group type, largely in favor of the assigned groups, but my study may be underpowered to capture these differences with certainty due to my sample size.

Moreover, I found that students may be indicating similar scores in a query item for substantially different reasons. For example, when queried regarding the benefits of their group, one of the top codes for both group types indicated that their laboratory experience was more enjoyable due to their group. However, I found a few subtle differences in the two group types for *why* it was enjoyable. For instance, an assigned student said, "*It was great to not have to choose people to be with all the time and I think our group got alone (along) really well,*" (Table 10). In contrast, a self-selected student who found *enjoyment* as the primary benefit of their group stated, *"It was pretty great to be able to work with people I already knew and could work well with."* These subtle but important differences suggest that at least some students in the self-selected sections grouped with friends and acquaintances with whom they already shared a level of comfort (Nhan and Nhan, 2019). Yet, this being a first-year introductory course, many students may not have already formed friendships with classmates and experiences may differ substantially within a single group type, or sometimes even within the same group. For instance, one self-selected student noted, *"...the other two in my group were close* *friends and I was not so it was sometimes hard to break into their discussion,"* (Table 7).

A possible limitation of my study is that the majority of the study population were White (77%) and women (66%), with substantial deviations from the mean within some sections, which limits capacity to consistently assign demographically heterogeneous groups. When possible I maintained a high gender distribution (Rosser, 1998), but it was not always possible, and ethnic categories were not considered in group assignments. As a result, implications will likely be different for classrooms that are more diverse. Studies have suggested that when left to their own devices, students will often choose to work with peers of the same ethnicity, gender, and similar academic ability (Mello 1993; Bacon *et al*, 1998; Johnson and Johnson, 2011; Freeman *et al*, 2017), which can cause lower cooperative outcomes due to homogeneity of perspectives. Environments with substantially higher ethnic or gender heterogeneity might be well served by accounting for demographic differences in group composition. Classrooms with more demographic diversity therefore should err on the side of assigned groups.

These uncertainties might matter in ways that are difficult to predict. The extensive peer group learning that students experience in their first year of college has broad and vital implications on social course objectives. The first-year college experience, including experiences in foundation courses such as the current course, serves as a place for students to build the social support needed to persist and succeed in their academic journey (Hausmann *et al*, 2007). In fact, many studies indicate it is often not just academic challenges but rather social factors such as in long-term lab groupings that influence student retention, performance, and sense of belonging (Tinto, 1987, 1993; Halpin, 1990; Napoli and Wortman, 1998). This is especially the case for students who are underrepresented, first-generation, lower in socio-economic background, who disproportionately leave STEM in the first two years (Graham *et al*, 2013). Therefore I cannot rule out that a more in-depth understanding of student group dynamics and attitudinal motives might reveal substantial factors which are important in balancing groups that improve the long-term student well-being beyond the course in study.

Implications

There is considerable debate in the literature on how groups should be formed (Donovan *et al*, 2018). Amongst these studies, some view that heterogeneous groups may be more effective for cooperative learning groups (Heller and Hollabaugh, 1991; Miller *et al*, 2012) yet others find similar outcomes with self-selected groups which are often more homogeneous (Lou *et al*, 1996; Baer, 2003; Jensen and Lawson, 2011). However, in my study, I did not find a substantial effect on student perceptions on cooperative group work from intervening in group heterogeneity based on students' prior knowledge and course interest. If there is an effect, my study suggests it is likely small. However, my study does indicate that students find substantial value in their experiences with group work in developing their academic and social skills. At the same time, students experienced diminishing concerns regarding group members over time. With such marginal differences, the debate around group composition may be a distraction, and the efforts at least relative to prior knowledge and course interest involved in balancing groups may serve to discourage instructors to little clear purpose. Rather, my data in combination with previous work suggests simply that more group work should be formed and more group work should be done.

Future Directions

My study has identified how students perceive their experiences and what they identified as benefits of assigned and self-selected group types. However, studies also show that student perceptions of their experiences influence their academic performance (Marsh *et al*, 1988; Ferriera and Santoso, 2008). Therefore in the next section of this work, I will examine any differences or alignments in academic performance for students in assigned and self-selected groups.

My study also suggests that both group types experience positive gains in cooperation from engaging in sustained group work. Yet while student perceptions and attitudes matter, it is worthwhile to explore how students learn and work with each other in their small peer groups; building evidence to inform practitioners on the effectiveness of grouping strategies and how students in small groups interact and learn.

V. COOPERATIVE LEARNING GROUPS IN A BIOLOGY LABORATORY COURSE: EXPLORING ELEMENTS OF COOPERATIVE LEARNING AND ACADEMIC PERFORMANCE (Study 2B)

Introduction

Learning in peer groups requires students to engage in discussions and interactions which benefit their learning and build life skills. When students work cooperatively in groups, they stimulate others to think as they explain the materials to each other (Leupen *et al,* 2020). Such positive group learning activities simultaneously benefit students sharing their understanding and other students in their groups (Chi, 2009; Chi and Wylie, 2014). Studies by Chi and Wylie (2014) and Leupen *et al,* (2020) found that interactive engagement in group learning produces far better learning outcomes than individual learning. These more recent findings are also backed up by decades of research by the Johnson brothers, who demonstrated that most activities even beyond the academic realm require far more cooperation in groups than competition (Johnson and Johnson, 1996, 1998, 2006). Cooperative learning therefore enhances student learning in knowledge acquisition, problem-solving ability, higher-level reasoning, and promotes student engagement and persistence (Johnson *et al,* 1998, 2006; Springer *et al,* 1999).

As discussed above and in Study 2A, there is consensus on the benefits of cooperative learning. However there is less consensus on how best to form the groups to maximize learning and improve students' group learning experience (Donovan *et al,* 2018; Nhan and Nhan, 2019). Therefore, our work below is directed toward demystifying the debate around how groups can be formed to enhance academic and social practices in the laboratory context.

In Study 2A, we examined how first-year students self-reported their experiences and what they identified as benefits and concerns of assigned base groups and self-selected group types while engaged in extended group work in their biology laboratory course. Assigned base groups are long-lasting and often academically heterogeneous with stable group memberships (Johnson and Johnson, 1991). On the other hand, the standard method for most group formation is to allow students to self-select their groups (observation, Chapman *et al,* 2006; Donovan *et al,* 2018; Nhan and Nhan, 2019). We concluded from Study 2A that students had similar perceptions on cooperative group work from group formation intervention. Overall, students in both group types indicated finding a tangible benefit in developing academic and social skills from their group experiences. We also found that over time as students continued to work with each other, their initial concerns related to aspects of group work, which included *unequal participation* and being *assigned a group grade* diminished.

In Study 2B, we expanded our study to explore how students in assigned and self-selected groups learn and work with each other in their small groups. To do so, we examined students from the same study population in both the assigned and self-selected groups. We explored the evidence of cooperative learning in both groups using the Cooperative Learning Observation Protocol (CLOP), a validated observation protocol (Kern *et al*, 2007), and we also evaluated their performance by examining their course grades. Along with our findings in Study 2A, utilization of the CLOP and evaluating their performance builds evidence to inform practitioners on the effectiveness of grouping strategies and more specifically how students in small groups interact and learn while engaged in a cooperative classroom task (Kern *et al*, 2007).

Literature Review

Cooperative learning stems from socio-constructivism and recognizes learning as a social phenomenon (Vygotsky, 1968). Cooperative learning involves purposely planned small groups of students working towards a shared learning goal (Deutsch, 1949a; Bybee *et al*, 2010). By doing so, students learn through their cooperative interactions with each other. The five essential elements of cooperative learning are positive interdependence, individual accountability, group processing, social skills, and face-to-face interactions (Johnson and Johnson 1998, 1999). In cooperative group work, the instructor provides guidance and support to the peer groups throughout the task, beginning with forming cooperative groups in a cooperative learning environment. In a successful cooperative group, students grow to see each other as strengths to their own ability to learn and succeed in the course. A positive perspective of group members is crucial because first-year students are often less trusting of each other's skills and knowledge (Finster, 1991). Instead, students can often see each other as competition, harming their and their peers' group experience. Cognitive studies demonstrate that many first-year students are at a dualistic position and over time move towards a relativistic position (Perry, 1999). Students with a dualistic perspective likely view instructors as the sole knowledge authority and do not yet see themselves or other students as equal contributors to the learning community (Kloss, 1994). A critical mass of students with a dualistic perspective can create a hostile group learning environment (Finster, 1991) which can be specifically harmful for underrepresented students in the STEM learning space

However, if students have a positive group experience, it can likely reduce the negative peer experience, bolstering their success in the course, encouraging them to feel a sense of belonging and promoting their learning (Beck and Malley, 1998; Zepke *et al*, 2006; Meeuwisse *et al*, 2018). Cooperative learning, as opposed to more competitive or individualistic methods, can create a sense of belonging and positively impact most students (Johnson and Johnson, 2005).

Cooperative Learning Observation Protocol: The CLOP is a validated mixed methods observation instrument to record and observe cooperative learning elements (Kern *et al*, 2007). The instrument is based on the CEPT-Core Evaluation Classroom Observation Program, an established and widely used instrument for recording cooperative learning elements in the secondary education environment (Lawrenz *et al,* 2002). The CLOP underwent four rounds of revision (Kern *et al*, 2007). The final version was observed by four separate groups to establish validity via the interrater reliability. The raters received a Cohen's Kappa of (κ)=.67, considered a *substantial agreement* (Landis and Koch, 1977).

The CLOP includes the incorporation of detailed field notes of the group learning activities. It also includes recording the frequency and evaluation of observed instances of cooperative learning. The researchers' purpose in developing the CLOP instrument was to evaluate the overall effectiveness of cooperative learning skills used by students engaging in cooperative tasks. As identified by Johnson and Johnson (1999) and encapsulated by CLOP, cooperative learning strategies include observations of a) positive interdependence: evidenced by team members moving towards the task's goal; b) group and individual accountability: evidenced by team members' active participation and holding others to the same standard; c) opportunities for group processing: evidenced by the team setting and reflecting on the task's goals and the process toward task completion; d) promotion of face-to-face interactions: evidenced by all ideas being listened to and valued by all group members. Another evidence for face-to-face interactions is when all group members contribute to the design and the task process. It also requires group members to help each other, listen to each other while being inclusive and constructively disagree. Finally, in developing the instrument, Kern *et al*, (2007), identified the importance of social skills in cooperative learning. Social skills are evident when members use proactive communication and generally agreed-upon social skills, including eye contact while talking to and respecting each other's ideas (Kern *et al*, 2007).

Research Questions

Our ***Research Questions*** for Study 2B are:

1) What, if any, are the elements of cooperative learning prevalent in self-selected and assigned groups in an introductory biology laboratory course?

2) How, if at all, does academic performance vary between students in the self-selected and assigned groups?

Methodology

Research Purpose and Context

The overarching purpose of this study is to explore the implementation of cooperative learning and examine any potential differences in academic performance for self-selected and instructor-assigned groups engaged in biology laboratory curriculum unit(s). To do so, the study employed the CLOP, a validated observational protocol (Kern *et al*, 2007) and evaluated students' overall course grades delineated by individual and group work. This study was at a liberal arts college situated in a midwest town (population ~11,000) in the United States of America. The participants are all in an introductory biology lab course in small groups of either three or four students, across four lab sections totaling 13-16 students each. In addition to the students, the labs consist of two undergraduate TAs and one instructor. The sections observed for the study were taught by two female instructors, each with more than five years of teaching experience in the given course.

The lab sections observed in the study were randomly assigned to either self-selected groups or instructor-assigned groups before the start of the semester. Instructor assigned groups were formed by cooperative learning methods (Johnson and Johnson, 2008), which include consideration of students' prior knowledge and interest in the subject, as described in Study 2A. All the students in the groups were first-year college students, aged 18 or above.

Curriculum Context

The small group interactions were observed during the 12th and 13th weeks of a 15 week course titled Organismal Biology. Organismal Biology is offered in the Spring semester sequentially following a course titled Principles of Biology. The primary group of students taking this course includes students with an intention to major in biology, biochemistry (with a pre-med focus), environmental studies, or the life science teaching majors. The course introduces students to the major organismal groups including bacteria, protists, fungi, plants, and animals. Students are introduced to these groups by examining distinguishing characteristics and the evolutionary relatedness among the organismal groups. The course closely explores organisms and their relationships to their environment, investigating how populations overcome environmental challenges through adaptations. Enrolled students attend four fifty-minute lectures a week and one three-hour lab for the duration of the ~fifteen-week semester. Students worked together beginning Week 1 of the labs.

Lab A: Trailblazing Termites: What sensory cues do termites use to follow trails?

Students were tasked with observing termites and developing and testing hypotheses regarding which sensory cues termites use to follow trails. Students first observed termite preference for colored pencils or pens as a small group, then worked as a class to determine which elements would be tested in the next round of experimental design. After completing the activity, students used statistical inference to determine the likelihood of the data supporting their hypothesis. Throughout the lab, students answered questions related to their observations and findings, as well as to the process of experimental design. As a part of the assignment, students also submitted a figure and an appropriate caption that summarized their results.

Lab B: Phylum Arthropoda: Anatomical diversity and cryptobiosis

In this activity, students explored reanimation from cryptobiosis in sea monkeys, a type of arthropod. Students began the class by designing a three-day experiment to explore the impact of environmental variables on reanimation from cryptobiosis using sea monkeys. After completing this, students worked as a group to explore arthropod body plans and identify differences in these morphological characteristics amongst the various subphyla. Post this activity, students cemented their understanding by completing an individual quiz. Once the group activity was completed, students wrote an abstract for their experiment and completed a computer-generated figure as a group.

Group Selection

Laboratory sections were randomly designated as instructor-assigned or self-selected groups at the beginning of the semester. Instructor-assigned groups were determined by evaluating students' prior knowledge through a pre-assessment (Appendix A), interest in the subject (Kern *et al*, 2007), and genders balanced across groups when possible (Rosser 1998; Handelsman *et al,* 2004). Interest in the subject was self-reported by students. Students were directly asked, *"Please rank your level of interest in this course,"* which included a Likert scale response ranging from *Strongly Interested* to *Strongly Disinterested*. The assessment questions were created by two faculty members who have taught the current and the preceding courses for over five years. A senior biology professor with extensive discipline-based education research (DBER) experience also verified the assessment. Gender was also self-reported by the students via an open-ended response.

Data Collection

CLOP Observations: As part of the CLOP, participants were recorded while engaged in the course. The team composition and the cooperative task at hand were also recorded. The CLOP instrument also includes the details of the instruction and lesson context leading to the task (Appendix B). Once the lab began, detailed field notes were taken, including the groups' activities and any notable interactions. The frequency of the interactions occurring in five-minute intervals of the observed cooperative learning activity was captured by the CLOP (Appendix C). As per a detailed rubric, high, medium, or low scale of the elements were coded (Table 11). Elements coded as high met all criteria defined by the CLOP, while medium elements met half of the criteria, or many at a minimal level, and low elements met few of the criteria, or most interactions that occurred were counterproductive to group processes. While the CLOP instrument, created by Kern, Moore, and Akillioğlu (2007), originated from observing a graduate-level course, the criteria effectively summarized the interactions observed in introductory biology laboratories, and no revisions were made to the rubric.

Six separate lab periods were observed for this study, while controlling for both the curriculum, instructor effect, and the students themselves (Fig. 10). Students in the assigned group, taught by Instructor 1, were observed during Labs A and B. Similarly, students in the self-selected group, taught by Instructor 2, were also observed in the same labs. Separate student groups were observed, both self-selected and assigned, taught by Instructors 1 and 2, respectively, to control for the impact of pedagogical differences experienced during varying teaching methods. This course lasted for the duration of the semester, and Lab A occurred in Week 13, while Lab B occurred in Week 12, allowing for the inference that students were comfortable with the expectations and structures of the course at the time of observation.

Table 11. Criteria for cooperative learning interactions in introductory biology lab activities. Modified from Kern *et al,* 2007.

Element	Description	Criteria
(P)ositive Interdependence	*There is evidence of group cohesiveness for accomplishing the task.*	*1.* All members acknowledge the task as a joint goal that they can achieve. *2.* Members undertake specific roles necessary to complete the task. *3.* There is established identity and ownership within the team. *4.* Members' actions are dependent on the actions of others. *5.* Each group member contributes their own unique resources to accomplish the task.
(I)ndividual Accountability	*Each individual group member takes responsibility for individual efforts and contributions towards the team.*	*1.* Each member individually participates and contributes to the team's effort to accomplish the task. *2.* All members are able to articulate and justify the process of group work. *3.* Each member contributes to team learning. *4.* The team takes responsibility to ensure each member understands the group process and can justify the output.
(G)roup Processing	*Students use ways to improve the processes team members use to maximize their own and each others' learning.*	*All members are involved in:* *1.* Giving feedback about the effectiveness of team processes. *2.* Setting goals related to accomplishing the task. *3.* Reflecting on the effectiveness of processes, on the goals of the task, on the progress made towards completing the task, and on accomplishing the task.
(F)ace-to-Face Promotive Interaction	*Students promote one another's success, through a supportive, encouraging, and praising environment.*	*1.* Team members give positive encouragement to each other for success and celebrate the team's work. *2.* All members value each others' ideas and efforts by respecting members. *3.* Each member feels safe to take risks during interaction without fear of peer judgement or retribution.

Grades: The final course scores were compared with grades obtained in the previous class. Final course scores were also compared while delineating students' performance in group assessments (project/exam scores) and individual assignments.

Data Analysis

All students' responses were downloaded to Microsoft Excel from Google Forms, and all information was de-identified. All incomplete data were removed, including a handful of students who directly enrolled in the course; likely via transfer credits from high school or from successful standing in standardized exams (i.e., Advanced Placement or college credit equivalent courses), thereby not completing the preceding course at this institution. Individual student scores were compared between the previous course and the current course using the Pearson correlation coefficient, also known as Pearson's r, to measure the linear association between the two scores. $r=1$ is a perfect correlation, and $r=-1$ means a perfect negative correlation. An r value of 0 indicates no correlation. A parametric independent t-test was run to compare grade items. The strength of the effect was determined using Cohen's d if a statistically significant difference was observed ($p<.05$). Individual student scores were also compared between assignments completed as a group and assignments completed alone using the Pearson correlation coefficient.

The study design and survey instruments were approved by the institution's IRB (approval #1819-0153).

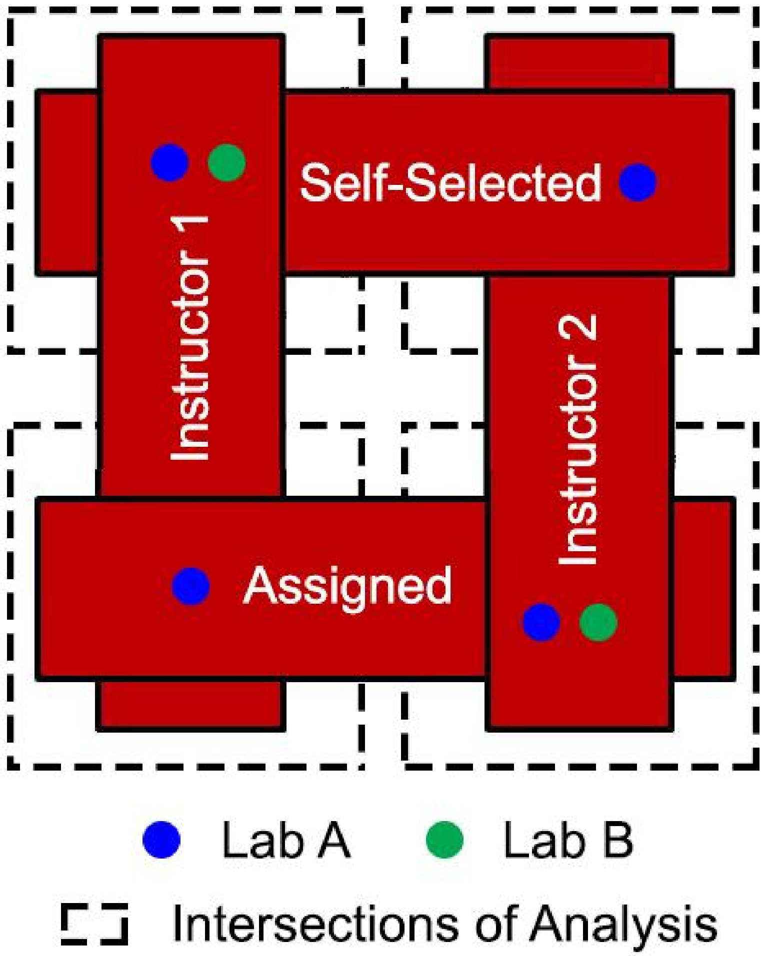


Figure 10: Areas of analysis for the CLOP. For one particular class module (Lab A), the CLOP protocol was implemented for all possible combinations of instructor and group type (each of the four corners). For another module (Lab B), analysis was confined to an Assigned group with Instructor 2 and a Self-Selected group with Instructor 1 (the lower right and upper left quadrants, respectively).

Results

Cooperative Learning Elements Analysis

<u>*Comparing assigned and self-selected for Instructor 1*</u>

Group Specifics: For Instructor 1, the assigned group was present in Section 1 of the course. We picked the group positioned closest to the video observation set-up for evidence of cooperative learning. Section 1 had a total of 15 students with a majority identifying as White (14). There was one Black student and the self-reported gender breakdown included nine females and six male students. The structure of the classroom allowed for groups to sit in rows or groups. Most students chose to sit in groups of four, facing each other, allowing for more face to face interaction between group members. The demographic information of the observed assigned group is recorded in Table 12.

The self-selected group was present in Section 2 of the course, and again, we observed the group closest to the video set-up. Section 2 also had a total of 15 students. 11 students self-reported as White, two as Asian, one as Asian and White, and finally one as unreported. The reported genders were four men, ten women and one as unreported. The structure of the classroom was the same as Section 1. Students opted to sit in groups of four, facing each other, to carry out their experiment. The demographic information of the self-selected group observed is recorded in Table 13.

Table 12. Instructor 1: Demographic information of the assigned group (Section 1).

	Year in School	Gender	Ethnicity
Student 1	1	Male	Black
Student 2	1	Female	White
Student 3	1	Female	White
Student 4	1	Male	White

Table 13. Instructor 1: Demographic information of the self-selected group (Section 2).

	Year in School	Gender	Ethnicity
Student 1	1	Female	White
Student 2	1	Male	Asian & White
Student 3	1	Female	Asian
Student 4	1	Female	White

Similarities: In Lab A, students observed termite behavior to hypothesize forms of sensory receptors used to follow trails. In doing so, students in both self-selected and assigned groups demonstrated all elements of cooperative learning for Instructor 1. In particular, we found positive interdependence and individual accountability elements more frequently in both (Fig. 11). For instance, self-selected and assigned students for Instructor 1 demonstrated high elements of individual accountability, evidenced by individual effort and contribution to the group (Kern *et al*, 2007). We observed all members in both groups articulating their ideas about controls and confounding variables to each other. Additionally, all students in the groups participated and asked each other, the instructor, and teaching assistants questions (Fig. 11).

Likewise, we also observed high elements of positive interdependence for both sections taught by Instructor 1. *Positive interdependence* is defined as group cohesiveness for accomplishing the task at hand (Table 11). In doing so, positive interdependence was observed in 15 out of 16 five-minute intervals (93.75%) for the assigned group. The element occurred throughout the class time with both explicit, such as direct communication, and implicit, such as nonverbal or subtle, components. In Lab A, positive interdependence was evident when members met the criteria of *acknowledging* *the tasks as a mutual goal*, *undertaking specific roles*, and *contribution of unique resources* (Table 11). For instance, in Instructor 1's assigned group, a student in the group had previously researched termites. He shared his knowledge with his group mates, discussing that termites prefer specific chemical compositions. His insight spurred a conversation amongst the whole group, and other members also contributed their knowledge about termites, experimental methods, and pen inks to the group discussion. The criterion of *undertaking specific roles* was evident when students prepared for experimental trials, while the *contribution of unique resources* occurred throughout the lab, as more discussion and interaction was needed to complete two separate instances of experimental design planning.

In the self-selected group, we observed positive interdependence in 15 out of 17 five-minute intervals (88.24%), as well. As before, this element of cooperative learning was met explicitly and implicitly as students communicated directly or observed the actions of their group. The self-selected group also demonstrated evidence of *undertaking specific roles* and *contributing unique resources.* Similar to the assigned group, students undertook roles necessary to complete the task as students prepared for experimental trials, commonly in an implicit manner. For example, one student noticed that a group member was collecting pens for the trails and decided to obtain termites for the group, while a different student prepared the cleaning supplies. All of this was done without explicit communication as students assessed the current needs of the task. As before, students met the criterion of *contribution of unique resources* throughout the lab, especially during instances of experimental design as students articulated their ideas and concerns.

Differences: We noted some small differences in the frequency of elements observed between the two group types for Instructor 1's sections. We observed more evidence of group processing, defined as improving the processes team members use to maximize their own and each others' learning (Table 11), in the assigned group, with ten intervals (five low, five medium) compared to seven (six low, one medium) in the self-selected group. The greater count of 'medium' elements suggests a diversity of group processing skills (Table 11). For instance, although we found frequent evidence of students giving feedback to each other on the experimental hypothesis and design in both groups, we saw more frequent feedback about the effectiveness of team processes in the assigned group. For example, students in the assigned group debated the best way to control confounding variables in their experimental design. Students reflected on the effectiveness of their controls, such as pen and pencil brands and colors, while questioning what confounding variables would exist with each design, such as the chemical composition of ink. To illustrate this interaction in the assigned group, Student 2 asked Student 4, *"But aren't they all different brands?"* to which Student 4 replied, *"They're all different brands, but they're all colored pencils."* This conversation highlighted how students were working through the components of group processing by reflecting on the effectiveness of their experimental design. While similar conversations were noted in the self-selected group, students failed to meet the criteria of *giving feedback about the effectiveness of team processes* as frequently as students in the assigned group, accounting for group processing being coded as 'medium' more frequently in the assigned group than the self-selected group.

In the assigned group, we observed frequent evidence of promotive interaction, defined as students supporting one another in a positive and encouraging environment (Table 11). Specifically, the criterion of *group members respecting and valuing the ideas and efforts of others* was met implicitly and explicitly with behaviors like nodding, verbal confirmations, and eye contact. Additionally, students in the assigned group demonstrated *feeling safe to take risks without fear of retribution or judgement,* as evidenced by frequent statements starting with *"I think...," "I said...,"* and *"What do you think about…".* Students were able to share their own thoughts and ideas with the group without facing negative feedback.

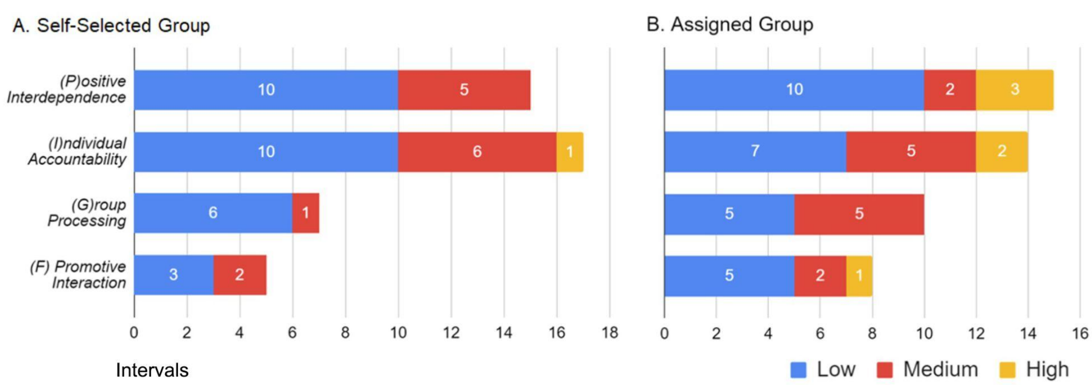


Figure 11: Incidence of cooperative learning elements in self-selected and assigned groups taught by Instructor 1 in Lab A. Students were observed during 5-minute intervals for a total of 18 intervals during the self-selected group's work time and 16 intervals during the assigned group's work time. Using the protocol, student interactions were coded as high, medium, or low. Elements coded as high met all criteria, while medium elements met half of the criteria, or many at a minimal level, and low elements met few of the criteria, or were counterproductive. Both groups engaged in positive interdependence and individual accountability the most frequently, but the assigned group engaged in more group processing and promotive interaction.

Comparing assigned and self-selected for Instructor 2

Group Specifics: For Instructor 2, the assigned group observed for cooperative learning elements was present in Section 3. Section 3 had 17 students, 14 of whom identified as female and three as males. The class was predominantly White (14) with three BIPOC students, each identified as Hispanic, Black, and Asian. The classroom structure allowed groups to sit in rows or groups, but the students chose to sit in a row of four and worked in groups of two instead of a whole group. We recorded the demographic information of Group 1 in Table 14.

The self-selected group was present in Section 4. The specific group observed for elements of cooperative learning and the entirety of Section 4 mostly consisted of White women. Section 4 had a total of 15 students. Ten students report as White, two as Hispanic, and one as Asian. These students self-reported as three males and ten females. The students opted to sit in rows of four and commonly subdivided themselves into two groups instead of one. For the self-selected group, we also found that all four students, while present in class, opted to work in pairs and rarely consulted across pair lines. We recorded the demographic information of the group observed in Table 15.

Table 14. Instructor 2: Demographic information of the assigned group (Section 3).

	Year in School	Gender	Race
Student 1	1	Female	White
Student 2	1	Female	White
Student 3	1	Male	White
Student 4	1	Female	Black

Table 15. Instructor 2: Demographic information of the self-selected group (Section 4).

	Year in School	Gender	Race
Student 1	1	Male	White
Student 2	1	Female	White
Student 3	1	Female	White
Student 4	1	Female	Asian

Differences: We found higher frequencies of cooperative learning elements in the assigned group than in the self-selected group for Instructor 2 (Fig. 12). In total, we observed 17 instances of cooperative learning in the self-selected group compared to 39 instances in the assigned group. The difference was chiefly for the group processing element, identified when students find ways to improve member processes to maximize learning for all members (Table 11). In the assigned group, we found three instances for group processing and none for the self-selected group. In the assigned group, we observed group processing, specifically for *reflecting on the effectiveness of steps taken towards the task*. Students in the assigned group, especially Students 3 and 4, had conversations about the experimental design and how effective it would be for testing termite sensory receptors. Specifically, the students interacted with their teaching assistant to confirm what confounding variables may be in their tentative design. After determining that their confounding variable would be factors like ink color, the students settled on a design that they believed would yield the best results, illustrating group processing.

Likewise, we also found little evidence of promotive interaction for self-selected students (one medium) but found several instances of the same for assigned groups for Instructor 2 (six low, two medium). In the assigned group, we observed criteria, including *students valuing each other's ideas and efforts*. One instance of this criterion occurred during the process of experimental design. Students were expressing their opinions about what constituted a control or a confounding variable, using language such as *"I think...,"* or *"I don't think..."* Students shared their opinions, and group members responded with verbal and nonverbal cues like agreement, nodding, and eye contact. Likewise, we also observed students *feeling safe to take risks without fear of retribution or judgmen*t. This criterion is closely related to feeling valued within the group, and students in the assigned group solicited the team's opinions by asking questions that led with "*What do you think…*" and sharing their own opinions with the group. These interactions indicate that students felt comfortable sharing their ideas without fearing negative responses.

Similarities: Elements of positive interdependence and individual accountability were frequent in both groups for Instructor 2. In the assigned group, we found 13 instances of positive interdependence (nine medium, four low), and we found seven instances (low) for the self-selected group (Fig. 11). For instance, in both the self-selected and assigned groups, we found that group members *acknowledged the task as a mutual goal* and *undertook specific roles*, both fulfilling the promotive interaction criteria. In addition, we also found evidence of group members *contributing unique resources* with fellow members in the assigned group.

We observed individual accountability frequently for both group types. In the self-selected group, we found nine instances (low), and in the assigned group, we found 15 instances (six medium, nine low). The self-selected students met the criterion of *individual participation and contribution* for every instance of individual accountability. For the assigned group, students met the same criterion while also engaging in *articulation and justification* and *contributions to team learning.* Yet, for both the assigned and self-selected groups, none of the elements were coded as high.

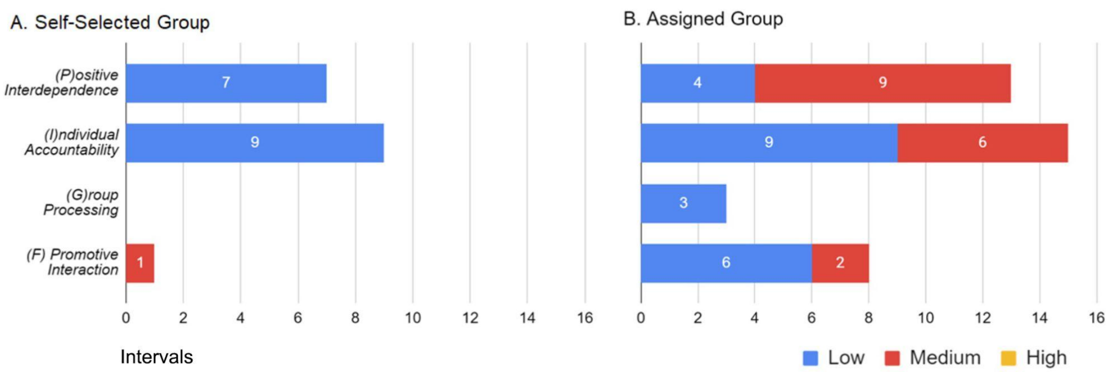


Figure 12: Incidence of cooperative learning elements in self-selected and assigned groups taught by Instructor 2 in Lab A. Students were observed during five-minute intervals for a total of 16 intervals both during the self-selected assigned groups' work time. Using the protocol, student interactions were coded as high, medium, or low. Elements coded as high met all criteria, while medium elements met half of the criteria, or many at a minimal level, and low elements met few of the criteria or were counterproductive. Both groups engaged in positive interdependence and individual accountability the most frequently, but the assigned group engaged in more group processing and promotive interaction.

<u>*Comparing instructional approaches*</u>

There were noteworthy differences in the pedagogical approach between the instructors even though the laboratory curriculum was identical. For instance, Instructor 1 opted to ask many questions to the individual groups about their experimental designs and in the process asked for definitions, and frequently pressed for students to justify their thoughts and answers. When asked questions by the students, Instructor 1 would answer briefly and then counter the student with another question, leading to the students processing their own questions and usually coming to a conclusion or answer themselves. The instructor also moved about the classroom, interacted with all groups, and only controlled the conversation during periods of instruction or whole-class brainstorming.

Instructor 2, in contrast, tended to address the entire class together, engaging in more direct teaching methods. For instance, Instructor 2 led a class discussion on the scientific process, detailing key elements of possible experimental designs, which the students subsequently voted on to adopt their experimental setup. Instructor 2's approach was notably more time efficient (the entire lab took 1 hour 45 minutes, as opposed to 2 hours and 24 minutes for Instructor 1), which decreased the amount of time in class for cooperative learning to occur but may have other benefits.

Lab B for assigned students (Instructor 1)

Along with individual accountability, positive interdependence was the most frequently observed cooperative learning element for the assigned groups in Lab B (Fig. 13). For positive interdependence, students commonly met the criteria of *undertaking specific roles* and *contribution of unique resources* during periods of experimental design, while evidence of *acknowledging a joint task* and *mutually dependent interactions* occurred during the experimental trial. For individual accountability, students commonly met the criterion of *individual participation and contribution*. Comparing Lab A and Lab B, group processing varied the most, likely due to the design of the activity. In Lab A, students carried out two instances of experimental design, one for students to test themselves and one to propose to the class, while Lab B had one instance of experimental design. However, even when accounting for differences in curriculum, we saw evidence of promotive interaction at similar frequencies between the two labs. Students met the criterion of *positive encouragement* through nonverbal cues such as eye contact and nodding, which fostered an encouraging environment and ultimately met the criterion of *feeling safe to take risks without fear of retribution or judgement*. Specific examples of these criteria and cooperative learning elements can be found in Appendix D.

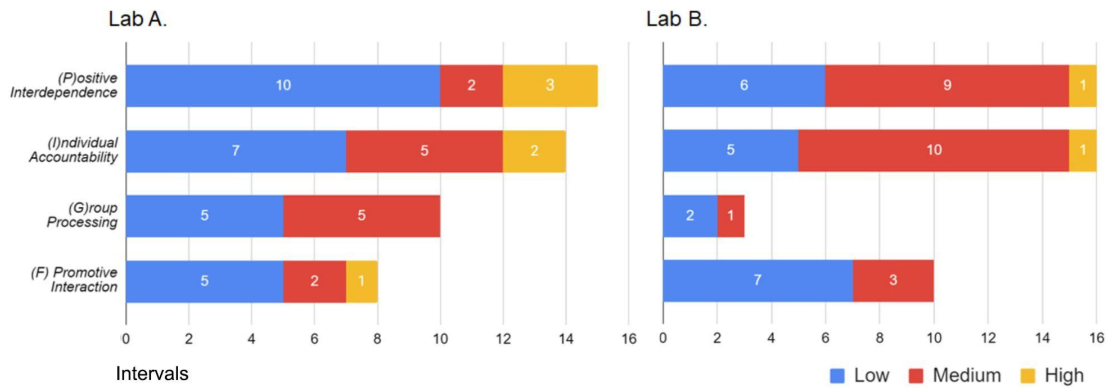


Figure 13: Incidence of cooperative learning elements in the assigned groups, comparing between lab modules. Students were observed during 5-minute intervals for a total of 16 intervals for both Lab A and Lab B. Using the protocol, student interactions were coded as high, medium, or low. Elements coded as high met all criteria, while medium elements met half of the criteria, or many at a minimal level, and low elements met few of the criteria or were counterproductive. Students engaged in all elements of cooperative learning, with positive interdependence and individual accountability occurring most frequently.

Lab B for self-selected students (Instructor 2)

During Lab B in Section 4, Student 2 was absent, and we observed Students 3 and 4, as Student 1 worked with another group. Positive interdependence and individual accountability were the most frequently observed elements of cooperative learning in Lab B for the self-selected group, and we saw little evidence of promotive interaction and no evidence of group processing (Fig. 14). For promotive independence, students commonly met the criteria of *acknowledging the task as a mutual goal* and *undertaking specific roles*, while we observed *individual participation and contribution* as the commonly met criterion for individual accountability. However, the frequency of cooperative learning elements differed between Lab A and Lab B, especially for positive interdependence (seven times and 12 times, respectively) and individual accountability (nine times and 17 times, respectively). These elements were still the most abundant evidence of cooperative learning in the self-selected group, and this variation may be attributed to limitations associated with the CLOP. In Lab A, we observed all four students in the self-selected group, which created strict application of the CLOP, as all four students needed to meet the criteria individually for the element to be coded at any level. Conversely, only two students were observed in Lab B, which lowers the threshold for coding elements. This difference likely affected the frequency of positive interdependence and individual accountability, but group processing and promotive interaction were not noted in either lab, regardless. Specific examples of cooperative learning for Lab B are outlined in Appendix E.

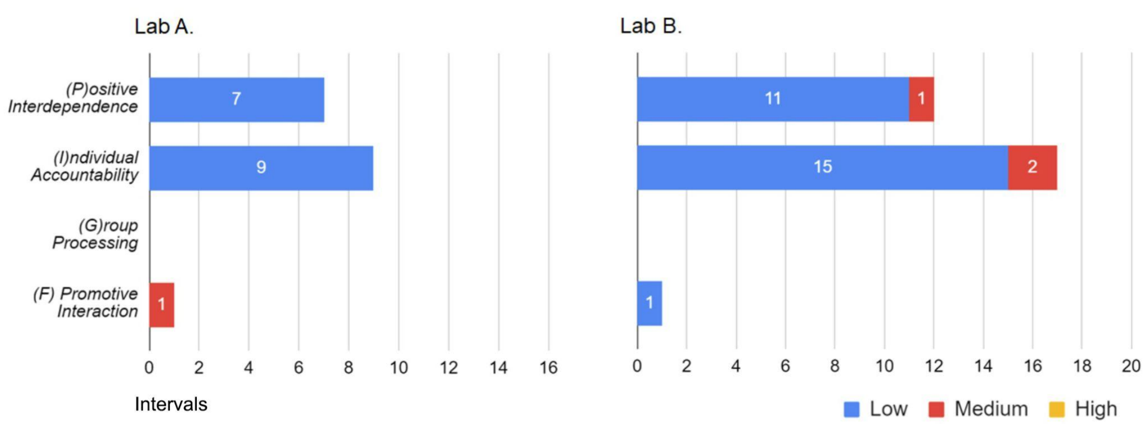


Figure 14: Incidence of cooperative learning elements in the self-selected group, comparing between lab modules. Students were observed during 5-minute intervals for a total of 16 intervals (x-axis) during Lab A and 21 intervals (x-axis) during Lab B. Using the protocol, student interactions were coded as high, medium or low. Elements coded as high met all criteria, while medium elements met half of the criteria, or many at a minimal level, and low elements met few of the criteria or were counterproductive. Students engaged in positive interdependence and individual accountability most frequently, with little to no evidence of group processing or promotive interaction. Notably, we observed four students for cooperative learning elements in Lab A, while only observing two students for cooperative learning elements in Lab B

Student Academic Performance in Assigned and Self-Selected Groups

Participants are representative of the students in the course

We found a statistically similar average score in the preceding course for students who participated in the study and students in the course at large, indicating our sample was representative (Appendix F). Similarly, we found no significant difference in the total course grades between non-participants and participants when comparing their grades in the current course (Appendix F).

Consistency of student scoring across sequential courses

There was a high level of correlation ($r=.831$, Pearson's r), without a notable difference between students from assigned and self-selected groups ($r=.856$ and $r=.814$, Pearson's r) when we compared all student scores between their current lecture course and the previous lecture course (Table 16). In comparing all student scores between the current lab course and the previous lab course, however, there was a positive but low level of correlation ($r=.286$, Pearson's r) (Table 17). Finally, students in assigned groups had a moderate correlation ($r=.364$, Pearson's r) to their previous lab course scores, somewhat higher than the low correlation in the self-selected group ($r=.217$, Pearson's r).

Table 16. Comparison of course lecture grades.

Population	Measure Lecture	Pearson Correlation Coefficient (*r*)
All Students	Previous vs. Current Lecture Course	.831
Assigned	Previous vs. Current Lecture Course	.856
Self-Selected	Previous vs. Current Lecture Course	.814

Table 17. Comparison of course lab grades.

Population	Measure (Lab)	Pearson Correlation Coefficient (*r*)
All Students	Previous vs. Current Lab Course	.286
Assigned	Previous vs. Current Lab Course	.364
Self-Selected	Previous vs. Current Lab Course	.217

Comparable academic performance between self-selected and assigned groups

We began the analysis by establishing a baseline from the participants' respective performances in the previous course. Assigned students on average scored 85.10, and self-selected students secured 84.19, with a p-value of .280 and a failure to reject the null hypothesis - or that is to say, in the earlier course, the students appeared to be similar performers. Likewise, students also had statistically similar academic performances in the current course overall, including lecture grades. For the current course the means between the groups indicated a p-value of .073, demonstrating a statistically similar score ($p>.05$) in academic performance for students in the assigned and self-selected groups (Table 18).

However, we found a statistical difference in academic performance when we dial down to see how students perform in the individual versus group tasks within the current course (Table 19). Laboratory grades completed in groups for 40 participants in the assigned group and 38 participants in the self-selected groups were analyzed using an independent t-test. Assigned group students ($M=88.73$, $SD=2.862$), compared to the self-selected group ($M=86.37$, $SD=3.22$) demonstrated significantly higher scores in laboratory group work, ($t(75)=3.394$, $p=.001$). Therefore group type may have a significant effect on group laboratory performance, with a moderate to large effect size (d=.77, Cohen's d), indicating that about three quarters of students from the self-selected group had scores below the mean of the assigned group. In contrast, there was no significant effect on laboratory grades of individually completed work ($t(74)=1.374$, $p=.174$), despite assigned groups ($M=87.81$, $SD=3.877$) attaining slightly higher scores than self-selected groups ($M=86.47$, $SD=4.621$) (Table 19).

Scores for individual assignments vs group assignments

We also found a high correlation between individual and group work assessment for self-selected students ($r=.501$, Pearson's r). In comparison we found a moderate correlation for assigned groups between their group and individual assignments ($r=.355$, Pearson's r). Therefore, assigned students had lower consistency between scores on the different types of assignments; students who scored highly on group work did not necessarily score highly on individual work, and vice versa. On the other hand, self-selected students had more similar scores on their individual and group completed work for the current course.

Table 18. Comparison of course grades between assigned and self-selected students.

Population	Measure	Mean	d.f.	Standard Error	*texp*	p-value
Assigned	Previous Course	85.10	29	1.088	.577	.28
Self-Selected		84.19	30	1.133		
Assigned	Current Course	84.79	39	.876	1.82	.073
Self-Selected		82.71	36	.910		

Table 19. Comparison of individual and group completed lab work between assigned and self-selected students.

Population	Measure	Mean	d.f.	Standard Error	*texp*	p-value
Assigned	Group Work	88.73	39	.453	3.394	.001
Self-Selected		86.37	36	.529		
Assigned	Individual Work	87.81	38	.621	1.374	.174
Self-Selected		86.47	36	.760		

Discussion

We observed many similarities between group types, but a greater frequency of cooperative learning elements in the assigned group when controlled for the instructor and the lab curriculum (Figs. 11, 12). In addition to the difference in cooperative learning experiences from group type, we also observed a practical but small difference in the scores of assignments completed as a group (Table 19).

Cooperative Learning in Self-Selected and Assigned Groups

Instructional approach is critical for cooperative learning

From our CLOP analysis, we found that although grouping strategy influences cooperative learning, the pedagogical approach of instructors likely has a greater influence. This is supported by the noticeable difference in the frequency of all cooperative learning elements between Instructors 1 and 2, regardless of group type (Figs. 11, 12). In particular, we found larger differences in group processing. We coded ten observations (five low, five medium) of group processing in the assigned group for Instructor 1, versus three (low) for Instructor 2. We also observed some differences in the promotive interaction related to greater high and medium criteria codings for Instructor 1, indicative of more diversity of interaction types (Fig. 11). We found eight observations (five low, two medium, and one high) of face-to-face promotive interactions for Instructor 1's assigned group in comparison to eight observations that were primarily low (six low, two medium).

These findings suggest that an instructor's pedagogical approach may play a critical role in the implementation of cooperative learning in the laboratory space. To state this more confidently, we would want to repeat this with other instructors and various types of laboratory curriculum. In our study, we found that group processing and promotive interaction are the two elements of cooperative learning most impacted (Figs. 11, 12). Group processing is particularly important for long-term groups and for culturally diverse classrooms (Tanner *et al,* 2003; Johnson and Johnson, 2005). Group processing exists when group members discuss how well they are achieving their goals and maintaining effective working relationships. Through group processing, students can develop the social skills necessary for interacting effectively with peers who come from different cultural backgrounds to their own (Johnson and Johnson, 2005). This will be relevant for other contexts where the study population is more culturally diverse than ours. During group processing, students must communicate with each other on how to effectively work as a group, working through points of tension and learning over time to cooperate more effectively with members.

We observed group processing in three of the four sections across both self-selected and assigned groups (Figs. 11, 12). In our study, the highest frequency of group processing was driven by instructional decisions that promoted students in small groups to learn from each other. To illustrate, Instructor 1 had students design experiments in small groups and present their proposed experiments to the class. This instructional approach encouraged students to share ideas and brainstorm together. As a result, group members had a vested interest in engaging in group processing because they would have to present their ideas to the class. Moreover, the experimental design that was most robust (i.e. all confounding variables, and details addressed, resources procured etc) was chosen by their peers as the whole class experiment. In this process, students were both learning from each other about the rigors of the scientific method but were also learning a social skill, which can translate to future group work. These instructional decisions resulted in more opportunities for group members to engage in group processing and promotive interaction.

Promotive interaction exists when students have time to converse verbally and discuss key concepts with each other (Johnson and Johnson, 2005). For promotive interaction, structured discussions are particularly helpful (Tanner *et al,* 2003). In the above example, for instance, Instructor 1 scaffolded the experimental design process with multiple stages for the instructor and different members to pose questions to each other, thereby stimulating an exchange of ideas amongst all group members. In doing so, students had the opportunity to verbally express their understanding to the whole group while promoting and encouraging each other's efforts towards the work at hand.

Because Instructor 1 chose to incorporate the experimental design stage at the group level, students were sharing ideas and brainstorming together, resulting in more opportunities to engage in group processing and promotive interaction. This activity did not occur in either of Instructor 2’s sections and likely as a result, contributed to lower cooperative learning but resulted in a temporally more succinct laboratory experience. While the benefits of shorter class times are not within the scope of this analysis, they should not be dismissed. However, the instructional decisions of the two instructors may have resulted in the variation of cooperative learning interactions observed (Figs. 11, 12).

There may also be some benefits to teaching to the whole class as engaged by Instructor 2. According to the developmental learning theorist William Perry, students early in their learning process are more likely to have a dualistic position, viewing instructors as the sole authority of knowledge and therefore may not yet be ready to learn heavily from their peers (Kloss, 1994). As such, instructional strategies that are grounded in social learning theories (i.e. group work), and other learning strategies geared toward the application of knowledge may be difficult for dualistic thinkers. Instead, according to Perry’s work, teachers should incorporate some features which are more conducive to a dualistic thinker, such as direct instruction, as students move further towards relativistic thinking (Myers, 2010). In this light, Instructor 2’s approach might be geared towards a gradual scaffolding approach, where students become more reliant on their peers over time. As the CLOP observations were during a small window of time (weeks 12 and 13), any progression over time would not have been captured.

The instructor difference sheds light on a potential reason why there remains a lack of consensus in the literature on group type formation (Donovan *et al,* 2018) - namely, the instructor implementing each study is a potential confounding variable. For instance, in multiple studies that examine grouping strategies, we found that instructor variation was not accounted for in examining cooperative learning (McInerney and Fink, 2003; Gaudet *et al*, 2010; Nhan and Nhan, 2019). While our study had an equal number of sections taught by both instructors for each group type, comparisons across different studies will not have this advantage. Further instructional differences could also explain some of the studies that suggest that self-selected groups may be more beneficial (Strong and Anderons, 1990; Bacon *et al*, 1999).

Curriculum influences the frequency of cooperative learning

Our CLOP analysis also suggests that there is likely some influence from the curriculum on cooperative learning in the class. For instance, there was a higher frequency of group processing in Lab A (week 13) compared to Lab B (week 12) for Instructor 1. We coded for three (two low, one medium) instances of group processing for Lab B, and ten (five low, five medium) for Lab A (Fig. 13). Our analysis suggests that the structure of the lab activities in part may contribute to the difference in frequency of group processing and other CLOP elements. For instance there were multiple brainstorming and feedback opportunities integrated into the lab activity resulting in group processing that were absent in Lab B.

However, we observed no consistent pattern in the change from modifying lab type, suggesting less influence from the curriculum compared to the differences from instructors (Fig. 13). For example, we observed similar patterns from modifying the lab curriculum type supported by increased elements of all CLOP measured elements for Lab B compared to Lab A when we controlled for instructors (Fig. 13, Appendix D). Additionally, when comparing between lab modules for Instructor 2 with the self-selected group, we noted higher levels in Lab B for positive interdependence (11 low, one medium) and individual accountability (15 low, two medium) versus Lab A (seven low and nine low, respectively), with other elements broadly similar (Fig. 14). On the other hand, the changes for Instructor 1 were minimal across the two lab types. (Fig. 13). While a direct comparison between group types or instructors for Lab B would be inappropriate with the present data, we found no evidence which contraindicated earlier interpretations between instructors or group types.

Academic Performances for Students in the Self-Selected and Assigned Groups

The high correlation between student scores across semesters in lecture classes (r=.831, Pearson's r, all students) reflects that traditional lecture classes utilize and reward a consistent set of skills. Lab classes, on the other hand, have a higher level of variability. This is seen in the lower correlation between lab scores across semesters (r=.286, Pearsons's r, all students), compared to the strong correlation found between lecture scores (r=.831, Pearson's r, all students) (Tables 16, 17). As a result, a laboratory experience can take a student who has strong academic performance in lecture and challenge them in ways that they may not usually face. On the other hand, students who may not be consistent strong performers in their previous courses have a chance to see their talents reflected in their course scores. This underlines the importance of group work at this stage of their major/college, especially as we see a huge drop off from STEM majors soon after the first two years (Chen, 2013).

While our academic performance data shows a significant difference in one type of assignment scores with a moderately strong effect size, we found that students in assigned and self-selected groups largely have similar academic performances reflected in their grades overall (Tables 18, 19). This finding is consistent with similar studies including the work of Harlow *et al,* (2007) who examined performance differences from team make-up which were instructor assigned (based on pre-assessment scores) and randomly assigned. They found no difference in performance from team make-up and subsequently recommended random selection of group members as it required less effort towards implementation.

When lab grades were delineated based on individual vs. group work, students in the assigned group secured significant gains beyond the self-selected group. The assigned group type had on average a relative gain of 2.36%, but only on assignments completed as a group. It is unclear if gains only in group and not in individual work indicate higher subject competence. One explanation might be, as suggested by Gibbs (2009), that group grading can potentially lessen student learning opportunities as some members may "freeload" from the effort of their peers, which might be more evident in academic heterogeneous groups due to larger potential gains for low scoring students - although our findings from Study 2A indicated that students report equal contributions from all group members. It is also possible that our study was underpowered to capture gains in individual work from assigned groups due to a limited sample size, though in any case our group type difference is small. Some previous studies suggest that heterogeneous groups have positive effects on individual learning (Heller and Hollabaugh, 1991; Handelsman *et al,* 2004). Our results are insufficient to support this claim, but are not in conflict with it based on our sample size.

As reported above, we also observed that self-selected students tended to obtain similar scores on their individual work and group work, (r=.501, Pearson's r). This might occur if self-selected students tended to complete their group work alone, supported by our analysis of the CLOP where multiple self-selected students segregated themselves away from table members during parts of the laboratory activities. Alternatively, it also suggests that self-selected groups were more academically homogeneous in membership, a finding supported by multiple previous studies (Mello, 1993; Bacon *et al,* 1998; Johnson and Johnson, 2011; Freeman *et al,* 2017). In contrast, the lower correlation between group and individual assignment for assigned students (r=.355, Pearson's r) is consistent with a positive effect from academic heterogeneous groups on their group assignments that did not translate to their individual assignments. However, any conclusions drawn from academic performance should be tempered with the understanding that academic outcome is a limited way to examine competency and to gauge student learning (Kohn, 1993; Farias *et al*, 2010).

Implications

Examining the evidence of cooperative learning elements in an introductory biology course allows practitioners to identify areas of growth and avenues for cooperative learning. Our findings in totality suggest that for instructors keen on cooperative groups, the instructor's pedagogical approach may be more vital than the grouping strategy. For instructors that are more aligned with cooperative learning, group type does not appear to matter as much. However, group type does appear to affect instructors who may be less familiar with cooperative learning.

Building on our findings for Study 2A our findings in Study 2B suggest structured group work requires scaffolded instructional support and an effective curriculum for reaping the benefits of cooperative learning. As a result, successful cooperation requires the fertile ecology of a pedagogical approach and a laboratory curriculum that supports cooperative groups. When these factors are considered we find there are more similarities in the evidence for cooperative learning regardless of group type (Fig. 11).

Once pedagogical approach and curriculum are controlled for, our study suggests there is evidence that assigned groups which are academically heterogenous allow for more cooperative learning opportunities for first year students (Figs. 11, 12). Additionally, our data, along with previous studies, suggest that this type of assigned group also allows for academic benefits, though limited to assessments from group work. However, before prioritizing grouping strategies, we recommend instructors 'practice' cooperative learning on small scales at shorter timelines while incorporating more scaffolding practices and making tacit education processes more explicit at the level of the curriculum. This is particularly vital for culturally diverse classrooms where students from minoritized backgrounds can be unfamiliar with STEM classroom cultures. Moreover, this strategy is also important for instructors gaining familiarity with cooperative learning.

Future Directions

In Study 2A and 2B of our study, we examined student perceptions and student practices in depth. Our findings were also supported with data on student performance while noting observations of pedagogical and curricular influences. In doing so, our study has identified that instructional approach and the curriculum are critical to the successful adoption of CL, likely even more than group type. Therefore a critical interrogation of the pedagogy and curriculum would help to complete our holistic examination of cooperative learning in this context, and allow me to develop a practical framework for implementation of cooperative learning in the first-year biology laboratory context.

VI. DISCUSSION

Below I briefly revisit my findings from the three chapters and then delve into the combined conclusions, implications, and future directions that demonstrate the evolution of my understanding of group learning practices in the STEM/biology context through an in-depth multi-faceted approach (Fig. 15).

	What	How	Why	Take-Away
1	Perceptions of and desire for active learning	Survey (one time point) (qualitative & quantitative)	To identify areas of improvement, examine student buy-in.	Students highly desired active learning, including chances to learn from peers.
2	Prior knowledge	Pre-semester assessment	To identify factors for heterogenization of groups.	Used to implement grouping strategy in Study 2A/2B.
2A	Perceptions of group work from two group types	Survey (two time points) (qualitative & quantitative)	To see attitude changes in response to exposure, investigate grouping strategies.	Students report academic/social benefits. Group types appear marginal, but qualitative coding suggests more positive experience in assigned groups.
2B	Cooperative behavioral practices in two group types	CLOP (controlling for instructor and curriculum)	To critically examine potential divergence from self-reported experiences.	More CL in assigned groups. Positive Interdependance and Individual Accountability are high. Group Processing and Promotive Interaction are lower, largely depending on instructor decisions.
2B	Impact of group formation strategy on performance	Grades (individual, group) (current & prior semester)	To see if heterogenization strategies have a measurable result on performance.	Assigned groups perform higher on group work. No difference for individual assignments.

Figure 15: Synthesized outline of data collected in my 3-study dissertation.

Chapter 3
Summary:

Study 1 of my dissertation revealed insights into common teaching practices across STEM classrooms from the students' perspectives. I identified how student experience compared to what students value as essential classroom practices. Specifically, I explored 23 teaching practices in graduate and undergraduate classrooms across three STEM colleges at a large mid-western university through survey-based research. I found that traditional lecturing was the most experienced practice in both graduate and undergraduate classes. Both graduate and undergraduate students overwhelmingly desired more time devoted to active learning than was experienced in their large STEM classes. Both these populations also wanted less lecturing as the primary mode of instruction. My study also demonstrated that no single active learning practice was universally preferred or unwanted, but students desired a variety of peer learning opportunities.

Chapter 4
Summary:

In Study 2A, I shifted focus to lab classes which are a critical part of the STEM learning experience and built for high levels of group work. Narrowing attention to the laboratory setting provided a richer data set by allowing us to examine how students experience working in groups as it is occurring. To do so, I investigated how students experience group work through two commonly identified methods of group formation, self-selected groups and students in sustained heterogeneous groups that were instructor assigned. I found that students had an improved perception of working in groups regardless of group types after the duration of a semester with continued group work. Specifically, students identified that working in groups provided both academic and social benefits. For instance, through an open-ended response, students stated that group members benefited them most by *contributing different perspectives,* bettering *their understanding* {of} *the material*, and making learning in the lab an *enjoyable experience*.

Study 2A also revealed that student concerns related to working in their groups diminished as they continued to work together over a semester.

Chapter 5
Summary:

In Study 2B, I continued to further my understanding from Studies 1 and 2A by examining the potential influence of group types and other factors for a successful cooperative group. Previously I investigated working in groups through the student perspective. In Study 2B, I utilized a different lens; via a validated observation protocol. I examined students from the same groups as Study 2A for evidence of cooperative learning in their small groups and investigated for any possible benefits towards their performance from group type.

There were similar patterns in cooperative learning across the two groups when I controlled for instructor pedagogical approach and laboratory activity. However, I found a greater frequency of cooperative learning elements observed in the assigned groups, suggesting that academically heterogenous assigned groups are better suited for cooperative learning. In addition, students previously cited the academic benefits of working in their groups. Therefore, I wanted to examine if working in their groups affected their performance. I observed a difference in the scores for group work favoring assigned groups (Table 19). However, these gains in performance as a measure of their academic performance did not transcend their group work. There was no significant difference in individual performance between self-selected and assigned groups (Table 19).

Conclusions

Across the three studies, I embarked on a journey to understand learning practices in the STEM/biology context through an extensive and multi-pronged approach (Fig. 15). Firstly I examined a broad population for their experiences in their respective STEM classrooms in Study 1. My results are supported by previous findings and point to systemic patterns also reflected in Studies 2A and 2B. I found there is student buy-in for active learning, and specifically, students valued more opportunities to learn from or via their peers. Therefore to examine student experiences of learning from peers, I shifted focus to the STEM lab, which has high levels of group work identified as the most common type of active learning, although limited in practice in the lecture space. My findings for Study 2A are consistent with Study 1 and highlight ways that group work can be enhanced and what students find valuable about working closely with their peers. My findings in Study 2B provide a different lens to analyze group work via a validated observational instrument and examining possible learning gains through course performance. I built on my previous findings of what students find valuable by analyzing how these group experiences manifest in the classroom. Finally, I provided strategies on how instructors can best critically analyze their curriculum and pedagogy to enhance the success of cooperative groups in their classrooms. By doing so, students can both maximize learning and increase the frequency of cooperative learning while working closely with their peers.

Limited Group Work in Lecture but there is Student Buy-in

Study 1 revealed that students had some experience but valued more group work type elements in their STEM classes. For instance, undergraduate students indicated they wanted more opportunities to study with their classmates outside of class (S7 Fig). Students were also open to working in assigned groups to complete their homework and other projects (S4 Fig). A close examination of the student responses revealed that students who currently experienced these things tended to want the same or more of these experiences (S4, S7 Figs). These findings are consistent with other studies when student attitudes towards learning in groups were evaluated (McInergney and Fink, 2003; Gaudet *et al,* 2010). For instance, the McIntergney and Fin (2003) study found that students who had undergone team-based learning found learning in groups improved their conceptual understanding of the course material.

My survey of all undergraduates also examined if students were receptive to being graded on group performance (S11 Fig). Students had concerns about being graded as a group, especially among those who had not experienced it. Apprehensions regarding group grading are also evident in previous studies with students perceiving an unfairness from members "freeloading" (Gibbs, 2009; Aivaloglou and Meulen, 2021). However, Study 1 suggests that students who experienced group grading in their STEM classrooms preferred a continuation of this practice (S11 Fig).

Students also wanted more opportunities to solve problems in groups while in class and were keen on think-pair-share, which students report as infrequently practiced (S13, S16 Figs). Think-pair-share is an active learning exercise where students respond to questions posed by the instructor after consulting with a classmate (Smith *et al*, 2009;

Tanner, 2009). Tanner (2009) suggests that active learning practices like think-pair-share may be a way for instructors to integrate evidence-based learning practices with minimal time and training commitment.

Collectively, Study 1 provided a broad survey of the STEM students across the undergraduate and graduate curricula with their experience and their desire for 23 teaching practices. Many of these are group-based learning practices, and this study, along with other studies (Ebert-May *et al*, 2011; Miller and Metz, 2014), has identified that although active learning is seldom practiced, the types of active learning most practiced are group-based learning (Cooper *et al*, 2017; Diesser *et al*, 2020). A focus on these suggests that students in lecture courses received limited experiences of working in groups. Moreover their perceptions are shaped by a variety of different environments including class type, class size, and instructor pedagogical choices.

Improved Attitudes and Diminishing Concerns over Time

My findings from Study 1, supported by Study 2A and previous work, indicate that students have some reservations about specific types of active learning, including facets of working with peers specifically related to assessment (Gibbs, 2009; Aivaloglou and Meulen, 2021). These apprehensions about group work are present in Study 2A's population as well (Table 7, S Table 3). First-year students in Study 2A revealed concerns regarding group work through a Likert scale and open-ended responses. Like the broad STEM population surveyed in Study 1, first-year biology students in Study 2A expressed concerns about being graded as a group (Table 8). In Study 2A, however, I witnessed an active decrease in all concerns across time. Specifically, trepidations related to being graded as a group were among the top three improvement areas across the semester; the most significant gain for assigned students and the third most for the self-selected students. This concern and others (i.e., unequal participation, communication issues) diminished as students continued to work together.

Study 2A also highlights that students improve their perceptions and attitudes to group work regardless of group type, supported by the qualitative and quantitative data. Students reported benefits including *contributing different perspectives*, better *understanding* (of) *the material*, and that working in their groups made labs a more *enjoyable experience*. Both types of groups also identified *learning to collaborate* and *gaining knowledge* as critical functions of their group. The qualitative and the quantitative data address similar positive gains from group work for both assigned and self-selected groups. This finding is supported by similar studies that examine grouping strategies (Donovan *et al*, 2018; Harlow *et al*, 2016). For instance, a study in a non-majors course placed low, medium, and high performing students in a variety of homogeneous and heterogeneous groups. Although they found more considerable learning gains for heterogeneous groups, all students across all groups had improved attitudes to working in groups (Donovan *et al*, 2018).

Consideration of Instructor Influence and Grouping Strategy on Cooperative Learning

Despite similar perceptions of working in groups for both group types, Study 2B's CLOP analysis and grade results showed some differences and revealed factors of influence previously unconsidered. I noted the vital role of an instructor's pedagogical approach for a learning environment adopting cooperative learning. Elements including positive interdependence and individual accountability arose from sustained group work regardless of the difference in instructor approach or group type. However, an instructor's pedagogical approach influences group processing and promotive interaction elements (Figs. 11, 12). Group processing manifests in members discussing how well they are achieving goals and maintaining effective group relations; both critical for long-term groups, especially if groups are culturally diverse (Tanner, 2003, Johnson and Johnson, 2005).

I found no measurable effect from group type on students' performance as measured by their end of the semester grade. This finding is consistent with Harlow *et al,* (2016), who did not find a difference from the team makeup on student learning. However, when I delineated grades between group and individual assessments, there was a statistical difference favoring assigned groups. I did not find this delineation in other studies. For instance, Harlow *et al* (2016) determined gains by evaluating final course exams. This finding and the lower correlation between group and individual assessment scores for assigned group students reflected that the assigned heterogeneous groups may benefit lower performing students, a conclusion also supported by other studies (Lou *et al*, 1996; Donovan *et al*, 2018).

Implications

My findings from Study 1 and other studies suggest that faculty should continue or expand active learning opportunities for their students across STEM (Freeman *et al*, 2014). In addition, the results from Study 2A reveal that regardless of grouping strategy, students experience substantial value from their experiences with group work in developing both their academic and social skills. Students also experienced diminishing concerns related to group members as they continued working together. Moreover, most of the benefits students reported from working in groups arose in both self-selected and assigned groups. My qualitative findings weakly suggested that grouping strategies lean toward assigned groups, and direct observations in Study 2B further supported this. These findings collectively advise that instructors will a) benefit from creating increased opportunities for students to work in groups b) maximize the student learning experience by creating stable and assigned member group tasks. Previous studies and my work recommends that assigned groups be academically heterogeneous while accounting for balancing genders across groups (Fig. 6) (Rosser, 1998; Handelsman *et al*, 2004). Study 2B suggests that successful group work is best accompanied by a scaffolded instructional approach and an effective curriculum (Fig. 16). In particular, instructional pedagogical choices that promote elements including group processing and promotive learning make the most notable impact for maximizing cooperative learning elements.

To illustrate, I have created a template based on the works of Johnson and Johnson (1998) for a possible laboratory exercise that incorporates cooperative learning from the beginning to the end of a single lesson in a first-year college biology lab (Appendix G). This template demonstrates how teachers can incorporate and identify all cooperative learning elements in their classrooms, with a pedagogical guide and a detailed curriculum for supporting students. It also provides strategies and examples for how instructors can encourage group processing at multiple levels, both within the group and the whole class. It demonstrates specific methods for how teachers can initiate positive interdependence, individual accountability, promotive interaction, and intergroup cooperation. In addition, I have also included methods of assessing and intervening in group dynamics when issues arise. This template is one example of how instructors can maximize the incorporation of CL to enhance learning while ensuring students reap experiences to build social skills from working in groups.

The concordance in student attitudes and concerns between the three studies implies that the studies' takeaways may be generalizable beyond lab courses. If instructors are to apply the takeaways to their lecture sections, my study is consistent with others to recommend utilizing group learning as an evidence-based active learning tool in lectures (Smith *et al*, 2009; Deslauriers *et al*, 2011; Tanner, 2017). Early in the semester, students will likely require reassurance and guidance given their self-reported apprehensions, especially regarding activities that result in group grades (Fig. 16). To circumvent this, assignments graded as a group can take place later in an academic semester after students have an opportunity to get to know each other and build trust. In the lecture space, instructors can also modify group sizes while maintaining heterogeneity. Studies by Smith *et al,* 2009 suggest that informal cooperative groups such as those formed during think-pair-share are an inroad to develop cooperative learning experiences in lectures which have similar benefits to assigned heterogeneous groups.

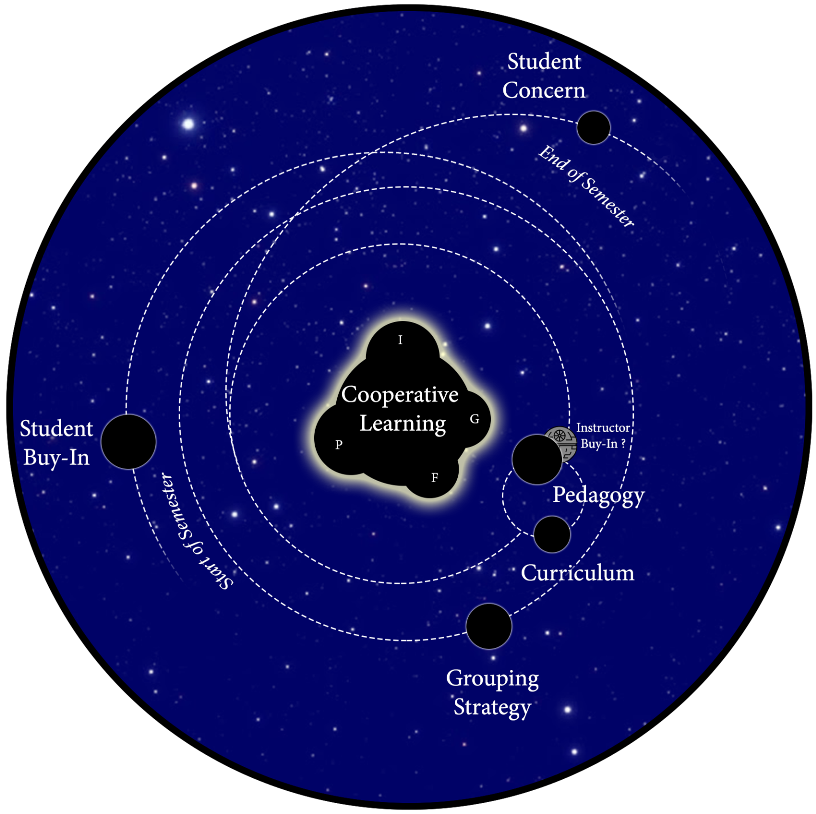


Figure 16: Proposed emerging framework for cooperative learning interactions. My proposed framework details a series of influences orbiting around Cooperative Learning practices. Student Concern is a potential factor with a close orbit at the beginning of the semester but moves to a less influential position over time. It is unclear if it is permanently left in a distant orbit, or returns in an elliptical orbit for the beginning of a new semester with a new group. Student Buy-In begins in a distant orbit not because it is unimportant overall, but because students reported a high interest to begin with - and therefore does not affect the model as strongly. However, the increase in Student Buy-In over time is reflected by a shift to a closer orbit, where it can enhance Cooperative Learning at the end of the semester. Grouping Strategy occupies a middle orbit, with secondary influences. However, it transits close enough to Curriculum and Pedagogy to interact with the pair. Curriculum and Pedagogy are a tightly interrelated binary pair with the highest influence on Cooperative Learning. These factors appear to influence Group Processing (G) and Promotive Interaction (F) the most, but have less influence on Positive Interdependence (P) and Individual Accountability (I). Instructor Buy-In was not directly measured in these studies, but evidence from other studies supports it as a shadowy presence that can affect Pedagogy and Curriculum.

Limitations

Though I have taken all possible care and consideration, limitations persist within these studies and must be considered. Regarding my study populations, the participants in this study were selected from an institution available to me (researcher) located in the Midwest at a predominantly white institution, potentially decreasing the extent that the findings could be generalized to all first-year STEM students across the nation.

Another important limitation of all three studies is that there was no opportunity to interrogate instructors for their perceptions of student learning practices or cooperative learning specifically. In the *Graduate and undergraduate-student perceptions of and preferences for teaching practices in STEM classrooms*, the study illustrated students' buy-in for more active learning opportunities, including learning in small groups with peers. The study met its goal to gauge students' self-reported experience and preferences and did not intend extension to faculty perceptions. However, adding faculty perceptions could identify areas of misalignment, if any, between students and faculty regarding classroom practices (Fig. 16). Gauging faculty insights will also reveal whether instructors buy into evidence-based teaching practices and how receptive they may be to future opportunities for training. Likewise, in the *Cooperative Learning Groups in a Biology Laboratory Course: Exploring Elements of Cooperative Learning and Academic Performance* study, I focused on recording and evaluating cooperative learning elements within groups (Kern *et al*, 2007), not specifically for instructor behavioral observations. However, instructional choices were captured through the CLOP, mainly when the instructors interacted directly within the groups. The absence of a full investigation of instructor behaviors and interactions hinders my understanding of the comprehensive instructional pedagogical influences that complement cooperative learning (Fig. 16).

Future Directions

My 3 study dissertation provides an emerging framework for supporting student learning and promoting positive experiences in the classroom while in small peer groups by examining from multiple perspectives (Fig. 16). My findings would be well-served by combining student perceptions with that of their instructors, thereby strengthening the framework. This additional data source will allow me to identify any areas of misalignment between student and faculty perspectives and understand faculty receptivity to potential further learning opportunities for effective CL implementation. Although I have evaluated the learning in different group types from multiple perspectives and developed several recommendations, the power of how these factors are wielded in concert ultimately depends on the instructor. Therefore as much as we have buy-in and participation from students, we also need the same mirrored from instructors. With faculty perspectives, I can further my framework for maximizing cooperation within and across STEM/biology peer learning groups (Fig. 16).

REFERENCES

Agrawal, R., Golshan, B., & Terzi, E. (2014, August). Grouping students in educational settings. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1017-1026).

Aivaloglou, E., & Meulen, A. V. D. (2021). An Empirical Study of Students' Perceptions on the Setup and Grading of Group Programming Assignments. *ACM Transactions on Computing Education (TOCE), 21*(3), 1-22.

Akiha, K., Brigham, E., Couch, B. A., Lewin, J., Stains, M., Stetzer, M. R., … Smith, M. K. (2018). What Types of Instructional Shifts Do Students Experience? Investigating Active Learning in Science, Technology, Engineering, and Math Classes Across Key Transition Points from Middle School to the University Level. *Frontiers in Education*, *2*, 68. https://doi.org/10.3389/feduc.2017.00068

Anderson, C. (2004). How much interaction is too much? *Usability Interface, 10*(3), 9-11.

Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers*. Jossey-Bass Publishers.

Astin, A. W., & Astin, H. S. (1992). Undergraduate Science Education: The Impact of Different College Environments on the Educational Pipeline in the Sciences. Final Report

Bacon, D. R., Stewart, K. A., & Silver, W. S. (1999). Lessons from the best and worst student team experiences: How a teacher can make the difference. *Journal of Management Education*, *23*(5), 467-488.

Bacon, D. R., Stewart, K. A., & Stewart-Belle, S. (1998). Exploring predictors of student team project performance. *Journal of Marketing education*, *20*(1), 63-71.

Bakeman, R., Gottman, J.M., 1986. Observing interaction: An introduction to sequential analysis. *Cambridge University Press*, New York.

Ballen, C. J., Wieman, C., Salehi, S., Searle, J. B., & Zamudio, K. R. (2017). Enhancing Diversity in Undergraduate Science: Self-Efficacy Drives Performance Gains with Active Learning. *CBE—Life Sciences Education*, *16*(4), ar56. https://doi.org/10.1187/cbe.16-12-0344

Barr, D. A., Gonzalez, M. E., & Wanat, S. F. (2008). The Leaky Pipeline: Factors Associated With Early Decline in Interest in Premedical Studies Among Underrepresented Minority Undergraduate Students. *Academic Medicine*, *83*(5), 503–511. https://doi.org/10.1097/ACM.0b013e31816bda1

Bazeley, P. (2013). *Qualitative data analysis: Practical strategies*. Sage.

Beichner, R., Bernold, L., Burniston, E., Dail, P., Felder, R., Gastineau, J., ... & Risley, J. (1999).Case study of the physics component of an integrated curriculum. *American Journal of Physics*, *67*(S1), S16-S24.

Beck, M., & Malley, J. (1998). A pedagogy of belonging. Reclaiming children and youth, 7(3), 133-137.

Bonwell, C. C., & Eison, J. A. (1991). *Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports*. ERIC Clearinghouse on Higher Education

Bonwell, C. C., & Eison, J. A. (1991). *Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports*. ERIC Clearinghouse on

Higher Education, The George Washington University, One Dupont Circle, Suite 630, Washington, DC 20036-1183.
Brame, C. (2016). Active learning. Vanderbilt University Center for Teaching.
Bransford, J. D., & Schwartz, D. L. (1999). Chapter 3: Rethinking transfer: A simple proposal with multiple implications. *Review of research in education*, *24*(1), 61-100.
Brazeal, K. R., Brown, T. L., & Couch, B. A. (2016). Characterizing student perceptions of and buy-in toward common formative assessment techniques. *CBE Life Sciences Education*, *15*(4). https://doi.org/10.1187/cbe.16-03-0133
Brewer, C. A., & Smith, D. (2011). Vision and change in undergraduate biology education: a call to action. *American Association for the Advancement of Science*, Washington, DC.
Brigati, J. (2018). Student Attitudes toward Active Learning vs. Lecture in Cell Biology Instruction. *The American Biology Teacher*, *80*(8), 584–591. https://doi.org/10.1525/abt.2018.80.8.584
Brown, A. L., & Palincsar, A. (1989). Guided, cooperative learning and individual knowledge acquisition in Resnick LB (ed) Knowing, learning and instruction: essays in honor of Robert Glaser.
Brown, T. L., Brazeal, K. R., & Couch, B. A. (2017). First-Year and Non-First-Year Student Expectations Regarding In-Class and Out-of-Class Learning Activities in Introductory Biology†. *Journal of Microbiology & Biology Education*, *18*(1). https://doi.org/10.1128/jmbe.v18i1.1241
Brownell, S. E., & Tanner, K. D. (2012). Barriers to faculty pedagogical change: Lack of training, time, incentives, and...tensions with professional identity? *CBE Life Sciences Education*, *11*(4), 339–346. https://doi.org/10.1187/cbe.12-09-0163
Bybee, R. W., Carlson-Powell, J., & Trowbridge, L. W. (2008). *Teaching secondary School science: Strategies for developing scientific literacy*. Columbus: Pearson/Merrill/Prentice Hall.
Campbell Jr, G. (1993). Visions of engineering education in century II. *The Porth Distinguished Lecture, University of Missouri at Rolla, MO*.
Carpenter, S. K., Witherby, A. E., & Tauber, S. K. (2020). On Students' (Mis)judgments of Learning and Teaching Effectiveness. *Journal of Applied Research in Memory and Cognition*, *9*(2), 137–151. https://doi.org/10.1016/j.jarmac.2019.12.009
Carrell, S. E., Fullerton, R. L., & West, J. E. (2009). Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics*, *27*(3), 439-464.
Cavanagh, A. J., Aragón, O. R., Chen, X., Couch, B. A., Durham, M. F., Bobrownicki, A., … Ledbetter, M. L. (2016). Student Buy-In to Active Learning in a College Science Course. *CBE—Life Sciences Education*, *15*(4), ar76. https://doi.org/10.1187/cbe.16-07-0212
Cavanagh, A. J., Chen, X., Bathgate, M., Frederick, J., Hanauer, D. I., & Graham, M. J. (2018). Trust, Growth Mindset, and Student Commitment to Active Learning in a College Science Course. *CBE—Life Sciences Education*, *17*(1), ar10. https://doi.org/10.1187/cbe.17-06-0107
Chapman, K.J.; Meuter, M.; Toy, D.; Wright, L. Can't We Pick our Own Groups? The Influence of Group Selection Method on Group Dynamics and Outcomes. *J. Manag. Educ.* 2006, 30, 557–569.

Cooper, K. M., Ashley, M., & Brownell, S. E. (2017a). A bridge to active learning: A summer bridge program helps students maximize their active-learning experiences and the active-learning experiences of others. *CBE—Life Sciences Education, 16*(1), ar17.

Corwin, L. A., Runyon, C. R., Ghanem, E., Sandy, M., Clark, G., Palmer, G. C., Reichler, S., Rodenbusch, S. E., & Dolan, E. L. (2018). Effects of Discovery, Iteration, and Collaboration in Laboratory Courses on Undergraduates' Research Career Intentions Fully Mediated by Student Ownership. *CBE—Life Sciences Education, 17*(2), ar20. https://doi.org/10.1187/cbe.17-07-0141

Charters, E. (2003). The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education: A Journal of Educational Research and Practice*, *12*(2).

Chen, X. (2013). STEM Attrition: College Students' Paths into and out of STEM Fields. Statistical Analysis Report. NCES 2014-001. *National Center for Education Statistics*.

Chi, M. T. H. (2009). Active-Constructive-Interactive: A Conceptual Framework for Differentiating Learning Activities. *Topics in Cognitive Science*, *1*(1), 73–105. https://doi.org/10.1111/j.1756-8765.2008.01005.x

Chi, M. T. H., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., … Yaghmourian, D. L. (2018). Translating the ICAP Theory of Cognitive Engagement Into Practice. *Cognitive Science*, *42*(6), 1777–1832. https://doi.org/10.1111/cogs.12626

Chi, M. T. H., & Wylie, R. (2014a). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Chi, M. T. H., & Wylie, R. (2014b). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Christie F., 2002, Classroom Discourse Analysis, London: Continuum.

Clark, V. L. P., & Creswell, J. W. (2014). Understanding research: A consumer's guide. Pearson Higher Ed.

Corwin, L. A., Graham, M. J., & Dolan, E. L. (2015). Modeling course-based Undergraduate research experiences: an agenda for future research and evaluation.*CBE—Life Sciences Education*, *14*(1), es1.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37-46.

Cooper, K. M., Ashley, M., & Brownell, S. E. (2017). Using Expectancy Value Theory as a Framework to Reduce Student Resistance to Active Learning: A Proof of Concept. *Journal of Microbiology & Biology Education*, *18*(2). https://doi.org/10.1128/jmbe.v18i2.1289

Cotner, S. H., Fall, B. A., Wick, S. M., Walker, J. D., & Baepler, P. M. (2008). Rapid Feedback Assessment Methods: Can We Improve Engagement and Preparation for Exams in Large-enrollment Courses? *Journal of Science Education and Technology*, *17*(5), 437–443. https://doi.org/10.1007/s10956-008-9112-8

Council, N. R. (2003). *BIO2010: Transforming Undergraduate Education for Future*

Research Biologists. Washington, DC: The National Academies Press. https://doi.org/10.17226/10497
Creed, T. (1986). Why We Lecture. In *Symposium: A Saint John's Faculty Journal* (Vol. 5, pp.17-32).
Creswell, J. W. (2009) Research design qualitative, quantitative and mixed methods approaches third edition.
Crossgrove, K., & Curran, K. L. (2008). Using clickers in nonmajors- and majors-level biology courses: student opinion, learning, and long-term retention of course material. *CBE Life Sciences Education*, *7*(1), 146–154. https://doi.org/10.1187/cbe.07-08-0060
Daniels, H. (2016). *Vygotsky and pedagogy*. Routledge. Vygotsky, L. S. (1980). *Mind in Society: The development of higher psychological processes*. Harvard university press.
DeMonbrun, M., Finelli, C. J., Prince, M., Borrego, M., Shekhar, P., Henderson, C., & Waters, C. (2017). Creating an Instrument to Measure Student Response to Instructional Practices. *Journal of Engineering Education*, *106*(2), 273–298. https://doi.org/10.1002/jee.20162
Denzin, N. K. (1978). The research act: A theoretical orientation to sociological methods (2nd ed.). New York: McGraw-Hill
Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(39), 19251–19257. https://doi.org/10.1073/pnas.1821936116
Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. science, 332(6031), 862-864.
Deutsch, M. (1949). An experimental study of the effects of co-operation and competition upon group process. *Human relations*, *2*(3), 199-23
Dewey, J. (1916). Education and democracy.
Dewey, J. (1966). Democracy and education (1916). *Jo Ann Boydston (ed.). The Middle Works of John Dewey*, *9*, 1899-1924.
Driessen, E. P., Knight, J. K., Smith, M. K., & Ballen, C. J. (2020). Demystifying the meaning of active learning in postsecondary biology education. *CBE—Life Sciences Education*, *19*(4), ar52.
Eberlein, T., Kampmeier, J., Minderhout, V., Moog, R. S., Platt, T., Varma‑Nelson, P., &
Ebert-May, D., Batzli, J., & Lim, H. (2003). Disciplinary research strategies for assessment of learning. *BioScience*, *53*(12), 1221-1228.
Ebert-May, D., Derting, T. L., Hodder, J., Momsen, J. L., Long, T. M., & Jardeleza, S. E. (2011). What We Say Is Not What We Do: Effective Evaluation of Faculty Professional Development Programs. BioScience, 61(7), 550-558. doi: 10.1525/bio.2011.61.7.9
Eberlein, T., Kampmeier, J., Minderhout, V., Moog, R. S., Platt, T., Varma-Nelson, P., & White, H. B. (2008). Pedagogies of engagement in science. *Biochemistry and Molecular Biology Education*, *36*(4), 262–273.https://doi.org/10.1002/bmb.20204
Eby, L. T., Hurst, C. S., & Butts, M. M. (2009). The redheaded stepchild in organizational

and social science research. *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences*, 219-246.

England, B. J., Brigati, J. R., & Schussler, E. E. (2017). Student anxiety in introductory biology classrooms: Perceptions about active learning and persistence in the major. *PLOS ONE*, *12*(8), e0182506.https://doi.org/10.1371/journal.pone.0182506

Esparza, D., Wagler, A. E., & Olimpo, J. T. (2020). Characterization of instructor and student behaviors in CURE and non-CURE learning environments: Impacts on student motivation, science identity development, and perceptions of the laboratory experience. CBE Life Sciences Education, 19(1). https://doi.org/10.1187/cbe.19-04-0082

Farias, G., Farias, C. M., & Fairfield, K. D. (2010). Teacher as judge or partner: The dilemma of grades versus learning. Journal of Education for Business, 85(6), 336-342.

Felder, R. M., & Brent, R. (2007). Cooperative learning. Active learning: Models from the analytical sciences, 970, 34-53.

Felder, R. M., Woods, D. R., Stice, J. E., & Rugarcia, A. (2000). The future of engineering education II. Teaching methods that work. *Chemical Engineering Education*, *34*(1), 26–39.

Ferreira, A., & Santoso, A. (2008). Do students' perceptions matter? A study of the effect of students' perceptions on academic performance. *Accounting & Finance*, *48*(2), 209-231.

Fink, L. D. (2013). *Creating significant learning experiences: An integrated approach to designing college courses*. John Wiley & Sons.

Finster, D. C. (1991). Developmental instruction: Part II. Application of the Perry model to general chemistry. Journal of Chemical Education, 68(9), 752.

Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(12), 6476–6483. https://doi.org/10.1073/pnas.1916903117

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Freeman, S., Theobald, R., Crowe, A. J., & Wenderoth, M. P. (2017). Likes attract: Students self-sort in a classroom by gender, demography, and Academic characteristics. *Active Learning in Higher Education*, *18*(2), 115-126.

Gaudet, A. D., Ramer, L. M., Nakonechny, J., Cragg, J. J., & Ramer, M. S. (2010). Small-group learning in an upper-level university biology class enhances academic performance and student attitudes toward group work. *PloS one*, *5*(12),e15821.

Gibbs, G. (2009). The assessment of group work: lessons from the literature. *Assessment Standards Knowledge Exchange*, 1-17.

Glaser, B.G., Strauss, A.L., 2009. The discovery of grounded theory: Strategies for qualitative research. *Transaction Publishers*.

Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science motivation questionnaire II: Validation with science majors and nonscience majors. *Journal of Research in Science Teaching*, *48*(10), 1159–1176. https://doi.org/10.1002/tea.20442

Gormally, C., Sullivan, C. S., & Szeinbaum, N. (2016). Uncovering Barriers to Teaching Assistants (TAs) Implementing Inquiry Teaching: Inconsistent Facilitation Techniques, Student Resistance, and Reluctance to Share Control over Learning with Students. *Journal of Microbiology & Biology Education*, *17*(2), 215–224. https://doi.org/10.1128/jmbe.v17i2.1038

Graham, M. J., Frederick, J., Byars-Winston, A., Hunter, A. B., & Handelsman, J. (2013). Increasing persistence of college students in STEM. *Science*, *341*(6153), 1455-1456.

Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2008). Increased Structure and Active Learning Reduce the Achievement Gap in Introductory Biology. *Science*, *332*(6034), 36(4):262–73. Retrieved from http://science.sciencemag.org/content/332/6034/1213.abstract

Halpin, R. (1990). An application of the Tinto model to the analysis of freshman persistence in a community college. *Community College Review*, 17(4), 22-33.

Harlow, J. J. B., Harrison, D. M., & Meyertholen, A. (2016). Effective student teams for collaborative learning in an introductory university physics course. *Physical Review Physics Education Research.*, 12, 010138. https://doi.org/10.1103/PhysRevPhysEducRes.12.010138

Handelsman, J., Ebert-May, D., & Beichner, R. (2004). Scientific teaching, (April). Retrieved from http://science.sciencemag.org/content/304/5670/521.short

Handelsman, Jo, Diane Ebert-May, Robert Beichner, Peter Bruns, Amy Chang, Robert DeHaan, Jim Gentile *et al*. "Scientific teaching." (2004): 521-522.

Heller, P., & Hollabaugh, M. (1991). Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups. *American Journal of Physics, 60*, 637–644.

Henderson, C., Khan, R., & Dancy, M. (2018). Will my student evaluations decrease if I adopt an active learning instructional strategy? *American Journal of Physics*, *86*(12), 934–942. https://doi.org/10.1119/1.5065907

Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science*, *326*(5958), 1410-1412.

Inagaki, K., & Hatano, G. (1987). Young children's spontaneous personification as analogy. *Child development*, 1013-1020.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative science quarterly*, *24*(4), 602-611.

Johnson, D. W., & Johnson, R. T. (1991). Cooperative learning and classroom and school climate.

Johnson, D. W., & Johnson, R. T. (1999). Making cooperative learning work. *Theory into practice*, *38*(2), 67-73.

Johnson, D. W., & Johnson, R. T. (2005). Cooperative learning, values, and culturally plural classrooms. *In Classroom Issues* (pp. 29-47). Routledge.

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story:

Social interdependence theory and cooperative learning. *Educational researcher*,*38*(5), 365-379.

Johnson, D. W., & Johnson, R. T. (2018). Cooperative learning: The foundation for active learning. *Active Learning—Beyond the Future.*

Johnson, D. W., Johnson, R. T., & Holubec, E. J. (1998). Cooperation in the Classroom: Interaction Book.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1991). Active learning: Cooperation in the college classroom. Edina, MN: *Interaction Book Company*

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1998). Cooperative learning returns to college: What evidence is there that it works? *Change, the magazine of higher learning*, 30(4),26–35. https://doi.org/10.10

Johnson, D. W., Johnson, R. T., & Smith, K. A. (2006). Active learning: Cooperation in the university classroom (3rd ed.). Edina, MN: *Interaction*.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (2014). Cooperative learning: Improving university instruction by basing practice on validated theory. *Journal of Excellence in College Teaching, 25*, 85–118.

Johnson, D. W., Johnson, R. T., Holubec, E. J., & Holubec, E. J. (1994). The new circles of learning: Cooperation in the classroom and school. *ASCD*.

Jones, M. H., Estell, D. B., Alexander, J. M., Jones, M. H., Estell, D. B., & Alexander, J. M. (2008). Friends, classmates, and self-regulated learning: discussions with peers inside and outside the classroom. *Metacognition Learning*, *3*, 1–15. https://doi.org/10.1007/s11409-007-9007-8

Jones, N. L., Peiffer, A. M., Lambros, A., Guthold, M., Johnson, A. D., Tytell, M., … Eldridge, J. C. (2010). Developing a problem-based learning (PBL) curriculum for professionalism and scientific integrity training for biomedical graduate students. *Journal of Medical Ethics*, *36*(10), 614–619. https://doi.org/10.1136/jme.2009.035220

Kaddoura, M. (2013). Think Pair Share : A teaching Learning Strategy to Enhance Students ' Critical Thinking. *Education Research Quarterly*, *36*(4), 3–24.

Karabenick, S.A., Woolley, M.E., Friedel, J.M., Ammon, B.V., Blazevski, J., Bonney, C.R., De Groot, E., Karacop, A., & Doymus, K. (2013). Effects of jigsaw cooperative learning and animation techniques on students' understanding of chemical bonding and their conceptions of the particulate nature of matter. *Journal of Science Education and Technology*, *22*(2), 186-203.

Keebaugh, A., Darrow, L., Tan, D., & Jamerson, H. (2009). Scaffolding the Science: Problem-Based Strategies for Teaching Interdisciplinary Undergraduate Research Methods. *International Journal of Teaching and Learning in Higher Education*, *21*(1), 118-126.

Kern, A. L., Moore, T. J., & Akillioglu, F. C. (2007, October). Cooperative learning: Developing an observation instrument for student interactions. In *2007 37th Annual Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports* (pp. T1D-1). IEEE.

Knight, J. K., Wise, S. B., & Sieke, S. (2016). Group random call can positively affect student in-class clicker discussions. *CBE—Life Sciences Education, 15*(4), ar56. https://doi.org/10.1187/cbe.16-02-0109 80/00091389809602629

Kohn, A. (1999). Punished by Rewards: The Trouble with Gold Stars, Incentive Plans,

A's, Praise, and Other Bribes. *Houghton Mifflin Harcourt*.
Landis, J. R. and Koch, G. G., "The measurement of observer agreement for categorical data," *Biometrics, 33*, 1977, 159-174.
Lane, A. K., Hardison, C., Simon, A., & Andrews, T. C. (2019). A model of the factors influencing teaching identity among life sciences doctoral students. *Journal of Research in Science Teaching*, *56*(2), 141-162.
Lawrenz, F, Huffman, D. Appeldoorn, K., and Sun, T., CETP Core Evaluation, Classroom Observation Handbook, Minneapolis, MN: *CAREI*, 2002.
Leupen, S. M., Kephart, K. L., & Hodges, L. C. (2020). Factors Influencing Quality of Team Discussion: Discourse Analysis in an Undergraduate Team-Based Learning Biology Course. *CBE—Life Sciences Education*, *19*(1), ar7.
Lopez, R. E., & Gross, N. A. (2008). Active learning for advanced students: The Center for Integrated Space Weather Modeling graduate summer school. *Advances in Space Research*, *42*(11), 1864–1868. https://doi.org/10.1016/J.ASR.2007.06.056
Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollnia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research, 66*,423–458.
Lunn, M., 1998. Applying k-sample tests to conditional probabilities for competing risks in a clinical trial. *Biometrics* 1662–1672.
Machemer, P. L., & Crawford, P. (2007). Student perceptions of active learning in a large cross-disciplinary classroom. *Active Learning in Higher Education*, *8*(1), 9–30. https://doi.org/10.1177/1469787407074008
Madson, L., Trafimow, D., & Gray, T. (2017). Faculty Members' Attitudes Predict Adoption of Interactive Engagement Methods. *The Journal of Faculty Development*, *Volume 31*(Number 3), 39-50(12). Retrieved from https://www.ingentaconnect.com/content/nfp/jfd/2017/00000031/00000003/art00005#Supp
Marsh, H. W., Byme, B. M. & Shavelson, R. J. (1988). A multifaceted academic self-concept: its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology, 80,* 366380.
Matz, R. L., Koester, B. P., Fiorini, S., Grom, G., Shepard, L., Stangor, C. G., Weiner, B., & McKay, T. A. (2017). Patterns of Gendered Performance Differences in Large Introductory Courses at Five Research Universities. *AERA Open, 3*(4), 233285841774375. https://doi.org/10.1177/2332858417743754
Mauri, M., Elli, T., Caviglia, G., Uboldi, G., & Azzi, M. (2017). RAWGraphs: a visualisation platform to create open outputs. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter* (p. 28). ACM.
Mayer, R. E., Stull, A., Deleeuw, K., Almeroth, K., Bimber, B., Chun, D., … Zhang, H. (2008). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes q. *Contemporary Educational Psychology*, *34*, 51–57. https://doi.org/10.1016/j.cedpsych.2008.04.002
Mcmillan, C., Loads, D., & Mcqueen, H. A. (2018). From students to scientists: The impact of interactive engagement in lectures. *New Directions in the Teaching of Physical Sciences*, *13*(1). https://doi.org/10.29311/ndtps.v0i13.2425
McInerney, M. J., & Fink, L. D. (2003). Team-based learning enhances long-term retention and critical thinking in an undergraduate microbial physiology course.

Microbiology Education, *4*(1), 3-12.
Meeuwisse, M., Severiens, S. E., & Born, M. P. (2010). Learning environment, interaction, sense of belonging and study success in ethnically diverse student groups. *Research in Higher Education, 51*(6), 528-545.
Mello, J. A. (1993). Improving individual member accountability in small workgroup settings. *Journal of management Education*, *17*(2), 253-259.
Michael, J. (2007). Faculty Perceptions About Barriers to Active Learning. *College Teaching*, *55*(2), 42–47. https://doi.org/10.3200/CTCH.55.2.42-47
Miller, C. J., & Metz, M. J. (2014). A comparison of professional-level faculty and student perceptions of active learning: its current use, effectiveness, and barriers. *Adv Physiol Educ*, *38*, 246–252. https://doi.org/10.1152/advan.00014.2014
Morris, J., & Chi, M. T. H. (2020). Improving teacher questioning in science using ICAP theory. *The Journal of Educational Research*, *113*(1), 1–12. https://doi.org/10.1080/00220671.2019.1709401
Morrison, C., & Williams. L. E. (1993). Minority engineering programs: A case for institutional support. *NACME Research Letter, 4*(1), 156–167.
Morse, J. M. (1994). "Emerging from the data": The cognitive processes of analysis in qualitative inquiry. In J. M. Morse (Ed.), *Critical issues in qualitative research methods* (pp. 22–43). Thousand Oaks, CA: Sage.
Myers, S. A. (2010). Using the Perry scheme to explore college student classroom participation. *Communication Research Reports*, *27*(2), 123-130.
National Research Council (2012). Discipline-based education research: Understanding and improving learning in undergraduate science and engineering.
Nhan, H., & Nhan, T. A. (2019). Different Grouping Strategies for Cooperative Learning in English Majored Seniors and Juniors at Can Tho University, Vietnam. *Education Sciences, 9*(1), 59.
Olson, S., & Riordan, D. G. (2012). Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics. Report to the President. *Executive Office of the President*. https://doi.org/10.1080/10668921003609210
Owens, D. C., Sadler, T. D., Barlow, A. T., & Smith-Walters, C. (2017). Student Motivation from and Resistance to Active Learning Rooted in Essential Science Practices. *Research in Science Education*, 1–25. https://doi.org/10.1007/s11165-017-9688-1
Patrick, L. (2020). Faculty and Student Perceptions of Active Learning. In *Active Learning in College Science* (pp. 889–907). Springer International Publishing. https://doi.org/10.1007/978-3-030-33600-4_55
Patrick, L., Howell, L. A., & Wischusen, E. W. (2016). *Perceptions of Active Learning between Faculty and Undergraduates: Differing Views among Departments*. *Journal of STEM Education: Innovations and Research* (Vol. 17). Retrieved from https://www.jstem.org/jstem/index.php/JSTEM/article/view/2121
Patrick, L., Howell, L. A., & Wischusen, E. W. (2018). Roles matter: Graduate student perceptions of active learning in the STEM courses they take and those they teach. *BioRxiv*, 502518. https://doi.org/10.1101/502518
Panitz, T. (1999). The motivational benefits of cooperative learning. *New directions for teaching and learning*, *78*, 59-67.

Patrick, O. (2013). Which way do we go in the teaching of biology? Concept Mapping, cooperative learning or learning cycle? *International journal of science and technology education research*, *4*(2), 18-29.

Perry Jr, W. G. (1999). *Forms of Intellectual and Ethical Development in the College Years: A Scheme. Jossey-Bass Higher and Adult Education Series*. Jossey-Bass Publishers, 350 Sansome St., San Francisco, CA 94104.

Pintrich, P. R. (1991). A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology*, *82*(1), 33.

Prince, M. (2004). Does active learning work? A review of the research. *Journal of engineering education*, *93*(3), 223-231.

Prince, M. (2004). Does Active Learning Work? A Review of the Research. *Journal of Engineering Education*, *93*(3), 223–231. https://doi.org/10.1002/j.2168-9830.2004.tb00809.x

Repice, M. D., Sawyer, R. K., Hogrebe, M. C., Brown, P. L., Luesse, S. B., Gealy, D. J., & Frey, R. F. (2016). Talking through the problems: A study of discourse in peer-led small groups. *Chemistry Education Research and Practice*, *17*(3), 555-568.

Rodrigo-Peiris, T., Xiang, L., & Cassone, V. M. (2018). A low-intensity, hybrid design between a “traditional” and a “course-based” research experience yields positive outcomes for science undergraduate freshmen and shows potential for large-scale application. *CBE Life Sciences Education, 17*(4). https://doi.org/10.1187/cbe.17-11-0248

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. Oxford university press.

Rewey, K. L., Dansereau, D. F., Dees, S. M., Skaggs, L. P., & Pitre, U. (1992). Scripted cooperation and knowledge map supplements: Effects on the recall of biological and statistical information. *The Journal of Experimental Education*, *60*(2), 93-107.

Ryan, G. W., & Bernard, H. R. (2003). Techniques to identify themes. *Field methods, 15*(1), 85-109.

Sato, B. K., Kadandale, P., He, W., Murata, P. M. N., Latif, Y., & Warschauer, M. (2014). Practice makes pretty good: Assessment of primary literature reading abilities across multiple large-enrollment biology laboratory courses. *CBE Life Sciences Education, 13*(4). https://doi.org/10.1187/cbe.14-02-0025

Saldaña, J. (2009). *The coding manual for qualitative researchers*. London ; Thousand Oaks, Calif.: Sage.

Sanders, D., & Welk, D. S. (2005). Strategies to scaffold student learning: Applying Vygotsky's zone of proximal development. *Nurse educator*, *30*(5), 203-207.

Sawyer, K., Frey, R., & Brown, P. (2013). Knowledge building discourse in peer-led team learning (PLTL) groups in first-year general chemistry. In *Productive multivocality in the analysis of group interactions* (pp. 191-204). Springer, Boston, MA.

Seidel, S. B., & Tanner, K. D. (2013). “What if students revolt?”—Considering Student

Resistance: Origins, Options, and Opportunities for Investigation. *CBE—Life Sciences Education*, *12*(4), 586–595. https://doi.org/10.1187/cbe-13-09-0190
Shin, J. C., & Harman, G. (2009). New challenges for higher education: global and Asia-Pacific perspectives. *Asia Pacific Education Review*, *10*(1), 1–13. https://doi.org/10.1007/s12564-009-9011-6
Sieber, S. D. (1973). The integration of fieldwork and survey methods. *American Journal of Sociology*, *78*(6), 1335-1359.
Silverthorn, D. U., Thorn, P. M., & Svinicki, M. D. (2006). It's difficult to change the way we teach: lessons from the Integrative Themes in Physiology curriculum module project. *Advances in Physiology Education*, *30*(4), 204–214. https://doi.org/10.1152/advan.00064.2006
Sims, C. (1992). What went wrong: Why programs failed. Science, 258(5085), 1185–1187 Singer, S., & Smith, K. A. (2013). Discipline‐based education research: Understanding and improving learning in undergraduate science and engineering. *Journal of Engineering Education*, *102*(4), 468-471
Singer, S. R., Nielsen, N. R., & Schweingruber, H. A. (2013). Biology education research: Lessons and future directions. *CBE—Life Sciences Education*, *12*(2), 129-132.
Slavin, R. E. (1980). Cooperative learning. *Review of educational research, 50*(2), 315-342.
Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of educational research, 57*(3), 293-336.
Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of educational research, 60*(3), 471-499.
Slavin, R. E. (1991). Synthesis of research of cooperative learning. *Educational leadership, 48*(5), 71-82.
Smith, C. V., & Cardaciotto, L. (2011). Is active learning like broccoli? Student perceptions of active learning in large lecture classes. *Journal of the Scholarship of Teaching and Learning* (Vol. 11). Retrieved from www.iupui.edu/~josotl
Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science*, *323*(5910), 122-124.
Smith, M. K., Wood, W. B., Krauter, K., & Knight, J. K. (2011). Combining peer discussion with instructor explanation increases student learning from in-class concept questions. *CBE—Life Sciences Education, 10*, 55–63
Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research, 69*, 21–51.
Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology
Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V, Cole, R., DeChenne-Peters, S. E., … Young, A. M. (2018). Anatomy of STEM teaching in North American universities. *Science*, *359*(6383), 1468 LP – 1470. Retrieved from http://science.sciencemag.org/content/359/6383/1468.abstract
Stone, D., Jarrett, C., Woodroffe, M., & Minocha, S. (2005). User Interface Design and Evaluation. San Francisco, CA: Morgan Kaufmann.

Stroebe, W. (2020). Basic and Applied Social Psychology Student Evaluations of Teaching Encourages Poor Teaching and Contributes to Grade Inflation:, A Theoretical and Empirical Analysis. https://doi.org/10.1080/01973533.2020.1756817

Tanner, K., Chatman, L. S., & Allen, D. (2003). Approaches to cell biology teaching: cooperative learning in the science classroom—beyond students working in groups. *Cell Biology Education*, *2*(1), 1-5.

Tanner, K. D. (2013b). Structure Matters: Twenty-One Teaching Strategies to Promote Student Engagement and Cultivate Classroom Equity. *CBE—Life Sciences Education*, *12*(3), 322–331. https://doi.org/10.1187/cbe.13-06-0115

Tanner KD (2009). Talking to learn: why biology students should be talking in classrooms and how to make it happen. *CBE Life Sci Educ 8*, 89-94.

Team, R. C. (2013). R: A language and environment for statistical computing. The George Washington University, One Dupont Circle, Suite 630, Washington, DC 20036-1183.

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Nicole Arroyo, E., Behling, S., …

Tinto, V. (1987). Leaving college: Rethinking the causes and cures of student attrition. Chicago: *The University of Chicago Press.*

Tinto, V. (1993). Leaving college: Rethinking the causes and cures of student attrition (2nd ed). Chicago: *The University of Chicago Press*

Trenor, J. M., Miller, M. K., & Gipson, K. G. (2011). Utilization of a think-aloud protocol to cognitively validate a survey instrument identifying social capital resources of engineering undergraduates. In *American Society for Engineering Education*. American Society for Engineering Education.

Tune, J. D., Sturek, M., & Basile, D. P. (2013a). Flipped classroom model improves graduate student performance in cardiovascular, respiratory, and renal physiology. *Advances in Physiology Education*, *37*(4), 316–320. https://doi.org/10.1152/advan.00091.2013

Tune, J. D., Sturek, M., & Basile, D. P. (2013b). Flipped classroom model improves graduate student performance in cardiovascular, respiratory, and renal physiology. *Advances in Physiology Education*, *37*(4), 316–320. https://doi.org/10.1152/advan.00091.2013

Tudge, J. R., & Winterhoff, P. A. (1993). Vygotsky, Piaget, and Bandura: Perspectives on the relations between the social world and cognitive development. *Human Development, 36*(2), 61-81.

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, *54*, 22–42. https://doi.org/10.1016/j.stueduc.2016.08.007

Velasco, J. B., Knedeisen, A., Xue, D., Vickrey, T. L., Abebe, M., & Stains, M. (2016). Characterizing instructional practices in the laboratory: The laboratory observation protocol for undergraduate STEM. *Journal of Chemical Education*, *93*(7), 1191-1203.

Vygotsky, L. S. (1968). Thought and language (newly revised, translated, and edited by Alex Kozulin).

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological*

processes. Harvard university press.
Walker, J. D., Cotner, S. H., Baepler, P. M., & Decker, M. D. (2008). Article A Delicate Balance: Integrating Active Learning into a Large Lecture Course Background and Course Design. https://doi.org/10.1187/cbe.08
Wankat, P. C., & Oreovicz, F. S. (1993). Models of cognitive development: Piaget and Perry. *Teaching engineering*, 264-283.
Wang, G., & Williamson, A. (2020). Course evaluation scores: valid measures for teaching effectiveness or rewards for lenient grading? *Teaching in Higher Education*. https://doi.org/10.1080/13562517.2020.1722992
Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. Springer.
Williams, M.A., and Wassenberg, D. 2017.Promoting Climate Change Literacy for Non-majors: Implementation of an atmospheric carbon dioxide modeling activity as an inquiry-based classroom activity. *CourseSource*. https://doi.org/10.24918/cs.2017.12
Willis, G.B., Royston, P., & Bercini, D. (1991). The use of verbal report methods in the development and testing of survey questions. *Applied Cognitive Psychology*, *5*, 251-267.
Witherspoon, E. B., Vincent-Ruz, P., & Schunn, C. D. (2019). When Making the Grade Isn't Enough: The Gendered Nature of Premed Science Course Attrition. *Educational Researcher*, *48*(4), 193–204. https://doi.org/10.3102/0013189X19840331
Xie, Y., Fang, M., & Shauman, K. (2015). STEM Education. *Annual Review of Sociology*, *41*(1), 331–357. https://doi.org/10.1146/annurev-soc-071312-145659
Zepke, N., Leach, L., & Prebble, T. (2006). Being learner-centered: One way to improve student retention? *Studies in Higher Education*, *31*(5), 587–600.

APPENDIX A
Pre-Semester Assessment

Pre-Assessment: We will begin with a few questions to gauge your prior knowledge of concepts you will learn in {*Current Course*}, as well as the retention of a few concepts from {*Previous Course*}. This pre-assessment *does not* in any way affect your {*Current Course*} grade. Just do the best that you can.

1. Which organisms are prokaryotes? (*circle all that apply*)
a. Fungi
b. Plants
c. Bacteria
d. Archea
e. Protists

2. Organisms that get their carbon from CO2 and then use light energy to make ATP are called:
a. Heterotrophs
b. Photoautotrophs
c. Chemolithotrophs
d. Heteroautotrophs

3. The way that plants move water from the roots up to the branches and then to the leaves is called:
a. Transpiration
b. Diffusion
c. Mass flow
d. Osmosis

4. The small pores on a plant leaf that the plant can open or close in order to diffuse gas into and out of the leaf are called:
a. Carboxylase
b. Phototrophins
c. Spores
d. Stomata

5. Which of the following animals are endothermic (*circle all that apply*)
a. Fish
b. Mammals
C. Amphibians
d. Birds

6. Hemoglobin is a very efficient carrier of:
a. Oxygen
b. Carbon Dioxide
c. Bicarbonate
d. Water

7. Mitochondria are responsible for ___.
a. Autophagy
b. Harvesting light energy to create sugars
c. converting glucose into ATP
d. all of the above

8. Which wavelength of light excites an electron to the highest energy state in the reaction center of Photosystem II?
a. Blue
b. Green
c. Red
d. Blue and red excite electrons to equivalent energy states

9. Alternative forms of the same gene are known as:

a. phenotypes
b. gametes
c. alleles
d. heterozygote

10. An early winter freeze decimates a frog population. The surviving population randomly happens to have larger spots on average than the initial population. If spot size is genetically determined, what event has the frog population experienced during the freeze?

a. Gene flow
b. Disruptive selection
c. Founder effect
d. Bottleneck effect

11. You are growing beans for the science fair and want to demonstrate disruptive selection. In these beans, you know that a single gene encodes for the color of the bean. R is the red allele, while W is the white allele. These alleles are codominant. Homozygous (RR) individuals produce red beans, while homozygous WW individuals produce white beans. Heterozygous (RW) individuals produce pinkish beans, due to different layers of the bean being different colors. Which of the following experiments would display disruptive selection?

a. Plant the beans in the garden and watch what happens to the frequency of the alleles.
b. Introduce a predator to the population, which only eats pink beans.
c. Distribute the beans in a field, and pick only the red and white beans while leaving the pink for the next generation.
d. Pick only the red beans

12. Suppose you are conducting an experiment that compares two species of unicellular eukaryotic organisms. You measure the pH of the mitochondrial matrix and the intermembrane space and the data are summarized in the table below. Between these two species of unicellular eukaryotic organisms, you will compare the rate of ATP synthesis. These differences in pH **do not** affect folding and functions of proteins within each organism.

Species	pH of mitochondrial matrix	pH of mitochondrial intermembrane space
Species A	7.8	6.9
Species B	7.8	5.9

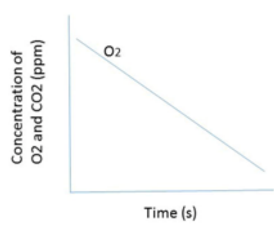
Which species will produce more ATP?

A. Species A
B. Species B
C. Species A and B will produce the same amount of ATP
D. pH is unrelated to ATP synthesis

13. When a person with type A blood has a child with an individual that has type B blood, what are the possible blood types of their offspring?

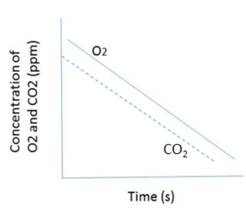
A. Type A
B. Type B
C. Type O
D. Type AB
E. All blood types are possible

14. You place a hissing cockroach in a sealed air-tight jar and place it in a water bath at 37℃. How would you expect the concentration CO_2 to change inside the sealed jar? I have drawn the change in O_2 in the figure below. Choose the figure that best represents your prediction; CO_2 lines are dashed, while O_2 lines are solid.

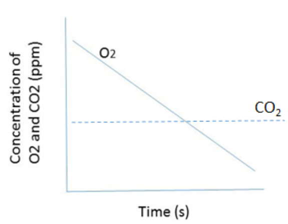


Circle one of the choices below that represents how CO_2 will change inside the sealed jar:

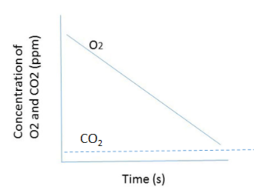
A.



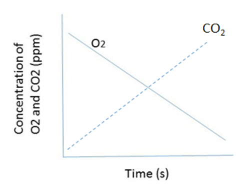
B.



C.



D.



APPENDIX B
Cooperative Learning Observation Protocol
Lesson Context
Kern et al, 2007

Section 1

Course/Level		**Date**	
Observer		**Instructor**	

Number of Students in Class		**Whole Class Demographic Information:**
Instructional Context:		

Group Specifics:

GROUP #/Name		**Seating Arrangement**			
Group Composition		**Length of class**			
Number of Students in Group		**Female**		**Male**	
Other:					

Section 2

Course/Level		**Date**	
Observer		**Instructor**	

Number of Students in Class		**Whole Class Demographic Information:**
Instructional Context:		

Group specifics:

GROUP #/Name		**Seating Arrangement**			
Group Composition		**Length of class**			
Number of Students in Group		**Female**		**Male**	
Other:					

Cooperative Task(s)

Cooperative Learning Elements (PIGS-Face) Descriptions

<table>
<tr><td colspan="4">(S)ocial Skills: Students display skills to promote a cooperative effort, leadership, decision-making, trust building, communication, and conflict management skills.

Notes: The Social Skills element encompasses the other four cooperative learning elements. This element focuses on the behavioral aspects of all four of the other elements. The Social Skills element is divided into four sub-elements that map to the other four elements: Forming, Functioning, Formulating, and Fermenting. Because this element was designed more as an implementation element rather than a single concrete, observable element in teamwork, we will build all of the behavioral pieces of Social Skills into the other fours elements of PIGS-Face.</td></tr>
<tr><td>(P)ositive Interdependence: There is evidence of group cohesiveness for accomplishing the task.</td><td>(I)ndividual Accountability: Individual group members take responsibility for individual efforts and contributions towards the team.</td><td>(G)roup Processing: Students use ways to improve the processes team members use to maximize their own and each other's learning.</td><td>(F) Promotive Interaction: Students promote one another's success, through a supportive, encouraging, and praising environment.</td></tr>
<tr><td>Examples:
▪ Roles as needed
o Facilitator, encourager, timekeeper, etc.
▪ Contributing unique background & skill</td><td>Examples:
Individual:
▪ Participation
▪ Contribution
▪ Engagement
▪ Ability to articulate & justify group procedures & results
Group:
▪ Group makes sure all understand task & procedures</td><td>Examples:
▪ Giving feedback to one another about effectiveness
▪ Setting goals or sub-goals
▪ Reflection on success
▪ Key words/sayings:
o "What we have so far"
o "Do we need to spend more time doing ___?"
o "Does everyone understand where we are?</td><td>Examples:
▪ Eye contact
▪ Name use
▪ Appropriate interruptions
▪ Suggestions by students respected (disagreement okay, but dealt w/ positively)
▪ Conflict is managed
▪ All members stimulate each other to draw out new ideas and fresh opinions
▪ Celebrate success</td></tr>
</table>

APPENDIX C
Cooperative Learning Observation Protocol
Element Rubric, *Kern et al, 2007*

Element	Low	Medium	High	NO
(P)ositive Interdependence *(There is evidence of group cohesiveness for accomplishing the task)*	Most of the High indicators are missing from the group interaction or are counterproductive to the group accomplishing the task.	Half of the High indicators are missing or many are attended to at a level that minimally meets the needs of accomplishing the tasks.	· All members acknowledge the task as a mutual/joint goal that they can achieve · Members create or undertake specific roles that the group needs to complete the task · Team has established identity and ownership · Members actions are dependent on the actions of other members · Each group member contributes their own unique set of resources (knowledge, experience) to accomplish the task.	Not observed
(I)ndividual Accountability *(Each individual group member takes responsibility for individual efforts and contributions towards the team.)*	Most of the High indicators are missing from the group interaction or are counterproductive to the individual accountability of the task.	Half of the High indicators are missing or many are attended to at a level that minimally meets the needs of individual accountability.	Each member individually participates in the task, and contributes to the team's effort to accomplish the task All members are able to articulate and justify the process of group work. Each member contributes to the team learning and to accomplishing the task. The team takes responsibility to ensure each member understands the group process and has the ability to justify the output.	Not observed
(G)roup Processing *(Students use ways to improve the processes team members use to maximize their own and each others learning.)*	Most of the High indicators are missing from the group interaction or are counterproductive to the improving group processing.	Half of the High indicators are missing or many are attended to at a level that minimally meets the needs of improving group processing.	All members are involved in: · giving feedback on the effectiveness of the team processes · setting goals related to accomplishing the task. · reflecting on the effectiveness, on the goals of the task, on the progress made, and on accomplishing the task.	Not observed
(F)Promotive Interaction *(Students promote one another's success, through a supportive, encouraging, and praising environment.)*	Most of the High indicators are missing from the group interaction or are counterproductive to promoting team and individual success.	Half of the High indicators are missing or many are attended to at a level that minimally meets the needs of promoting team and individual success.	· Team members give positive encouragement to each other for success and celebrate team's work. · All members value each others ideas and efforts. · Each team member feels safe to take risks during interaction without fear of peer judgment or retribution	Not observed

APPENDIX C
Cooperative Learning Observation Protocol Template

Kern et al, 2007

Interval	Element					Notes	Quotes
		L	***M***	***H***	***N***		
0-5	***P***						
	I						
	G						
	F						
5-10	***P***						
	I						
	G						
	F						
10-15	***P***						
	I						
	G						
	F						
15-20	***P***						
	I						
	G						
	F						
20-25	***P***						
	I						

	G						
	F						
Total PIGF Interactions		**L**	**M**	**H**	**N**		
		0	**0**	**0**	**0**		
Frequency of PIGF	***P***	**0**					
	I	**0**					
	G	**0**					
	F	**0**					
Observed Patterns							
Comparison: Section 9 versus Section 10							

APPENDIX D

Lab B for Assigned Students (Instructor 1)

Cooperative learning in Lab B: Along with individual accountability, positive interdependence was the most frequently observed cooperative learning element for the assigned groups in Lab B (Fig. 13). We observed positive interdependence 16 times in 16 5-minute intervals (100%). These interactions occurred consistently in each time interval, and many criteria were met implicitly. All criteria of positive interdependence were met at some point during the lab, but *group members undertaking specific roles* and *contribution of unique resources* were the most consistently met.

For example, the criterion of *contribution of unique resources* was met as students discussed arthropod morphological composition. As students were categorizing the variety of arthropod specimens by morphology, Student 1 shared that a specific body composition was characteristic of one group, while Student 2 mentioned the significance of different body structures. Additionally, students frequently met the criterion of *undertaking specific roles* implicitly. Students divided up work depending on the needs of the current task, but occasionally, students would direct one another or announce their roles. For instance, Student 2 communicated obtaining sea monkey cysts while Student 3 prepared the containers. Even with nonverbal interactions, the team met the criteria for positive interdependence and productively worked towards task completion. These criteria, *undertaking specific roles* and *contribution of unique resources*, were commonly met during periods of experimental design, while criteria such as *acknowledging a task as a mutual goal* and *mutually dependent interactions* occurred during the experimental trials.

We also frequently observed individual accountability during both lab activities (Fig. 13). In Lab B, we observed individual accountability 16 times (100%), coded as low five times, medium ten times, and high once. Group members frequently met the criterion of *individual participation and contribution.* All students would be actively working on the task at hand, whether through collecting supplies or recording observations, and contributing to answering assignment questions. Additionally, members met the criterion of *articulation and justification*, especially during periods of experimental design. For example, Student 1 was listing environmental variables that could impact reanimation out of cryptobiosis, and Student 2 responded "*Shouldn't we be testing for one thing?*" Student 1 then explained that she was listing variables so that the group could choose what they thought was best. These interactions indicate that students were able to justify their actions, even when countered by other group members.

Cooperative learning differences between Lab A and B: We observed group processing in the assigned group, but the element varied the most between Lab A and Lab B. In Lab B, group processing was observed only three times (18.75%), especially at the beginning of the class as students created their experimental design. The element was coded as 'low' (2) and 'medium' (1), and students met the criterion of *reflection on the effectiveness of group processes* the most frequently. Conversely, group processing was observed ten times in Lab A (62.5%), and the element was coded as 'low' and 'medium' five times each. There was more group processing during the beginning of the class, as Lab A required students to create their experimental designs as a group and as a class. The curriculum variation between Lab A and B is likely responsible for the frequency difference in the element between the two lab activities. In Lab B, students created an experimental design for their small groups, and they were tasked with comparing termite responses to pens and colored pencils. Then, students created a second experimental design to share with the class, primarily focusing on pens to test chemoreceptors or photoreceptors. Conversely, Lab A had one instance of experimental design, when students created an experiment to test the impact of environmental factors on arthropods' reanimation from cryptobiosis.

Cooperative learning similarities between Lab A and B: We observed promotive interaction consistently between the two lab activities. In Lab B, the promotive interaction element occurred ten times (62.5%), was coded as 'low' (7) and 'medium' (3). There was not a consistent pattern of promotive interaction throughout the class time, but students appeared to meet the criteria of the element after completing experiments or during times of discussion. While we rarely observed *positive encouragement* in verbal statements, students did participate in nonverbal encouragement with head nods, engaged body language, and name usage. In one instance, students were preparing the experimental dish with tape, and Student 2 was struggling with the task. Student 1 gave positive verbal encouragement, saying "*I believe in you, [Student 2],"* to which Student 2 responded "*This is kinda fun, I enjoy this. This is teamwork right here.*" This interaction meets the criterion of *positive encouragement*, but also illustrates a positive environment where students meet the criterion of *feeling safe to take risks without fear of retribution or judgment.* Similar patterns occurred in Lab A, where students engaged in promotive interaction eight times (50%). The element was coded as 'low' (5), 'medium' (2), and 'high' (1). However, the frequency of promotive interaction increased at the beginning of class as students developed experimental designs and collaborated as a group. Students often met the criteria of *group members respecting and valuing the ideas and efforts of others*. Examples include behaviors such as nodding, verbal confirmation, and eye contact among group members. Students also met the criterion of *feeling safe to take risks without fear of retribution or judgement* by comfortably sharing their own ideas with the group, demonstrated by statements such as "*I think…" and "I said…"*

APPENDIX E

Lab B for Self Selected Students (Instructor 2).

Cooperative learning in Lab B: During Lab B, Student 2 was absent, and we observed Students 3 and 4, as Student 1 worked with another group. Positive interdependence and individual accountability were the most frequently observed elements of cooperative learning in Lab B (Fig. 14). During Lab B, positive interdependence was observed 12 times (57.15%) in 21 5-minute intervals and was coded as 'low' (11) and 'medium' (1). Students commonly met the criteria of *acknowledging the task as a mutual goal* and *undertaking specific roles*. For example, students sought feedback from their teaching assistant about the efficacy of their experimental design and the best way to represent their results, acknowledging that they understood the goal of the lab. Additionally, students implicitly took on unique roles to contribute to experimental trials, such as Student 3 obtaining supplies and Student 4 preparing the Petri dishes. These criteria were met during periods of experimental design and setup, but in general, the students relied heavily on the instructor's and teaching assistants' guidance.

Additionally, we observed individual accountability 17 times (80.95%), with the majority of instances being coded as 'low' (15). Students primarily met the criteria of *individual participation and contribution*, but rarely met any of the other criteria. Many interactions associated with individual accountability were counterproductive to group processes, such as students actively admitting they were not paying attention to instructor directions or articulating their desires to efficiently accomplish tasks in order to leave. However, students did meet *articulation and justification* of their ideas, especially when classifying the arthropods by morphology. For example, Student 2 articulated a specific feature of an arthropod phylum to Student 1, which demonstrated evidence of cooperative learning.

Cooperative learning differences between Lab A and B: The frequency of cooperative learning elements differed between Lab A and Lab B, especially for positive interdependence and cooperative learning. During Lab B, we observed positive interdependence 12 times (11 low, 1 medium) and individual accountability 17 times (15 low, 2 medium). Conversely, in Lab A, we observed positive interdependence 7 times (low) and individual accountability 9 times (low). This variation may be attributed to differences in activity content and structure. For example, Lab A required more brainstorming of experimental designs, as students first had to create a design for their small group then for the broader class, while Lab B restricted which variables students could test, such as temperature and water salinity, potentially limiting students abilities to participate in group processing or promotive interaction. However, while this hypothesis is supported by the findings for Instructor 1, it does not explain the variation captured in these observations. Therefore, some variation may be due to elements not captured by the CLOP. During Lab A, we observed all four members of the self-selected group, which demands that all four members meet the criteria of each element of cooperative learning. During Lab B, only Students 3 and 4 were observed, which lowered the proportion of students needing to engage in cooperative learning for elements to be coded as low, medium, or high.

Cooperative learning similarities between Lab A and B: While the frequency of positive interdependence and individual accountability varied between Labs A and B, we still noted these elements of cooperative learning as being the most abundant in the self-selected group. Students also met similar criteria, with the most common being *acknowledging the task as a mutual goal* and *undertaking specific roles* for positive interdependence and *individual participation and contribution* for individual accountability in both Lab A and Lab B. For example, students undertook specific roles in both labs in order to complete the activity. In Lab A, students divided roles amongst themselves into categories such as termite observer, supply manager for pens, pencils, and termites, and experimental trial timer, while in Lab B, these tasks included collecting different water samples and obtaining arthropod cysts. Students also consistently participated in both labs by offering ideas, suggestions for design, and completing questions in their notebooks. Additionally, we rarely observed group processing and promotive interaction in both labs.

APPENDIX F

Participants are representative of the students in the course

We found statistically similar average scores in the preceding course for students who participated in the study and students in the course at large, indicating our sample was representative. To establish this baseline, we ran a parametric independent t-test to verify that students who participated in the study were representative of the class as a whole for their grades in the current and previous courses. We compared students' course grades, including both their lecture and lab scores. Because final grades were recorded as a letter grade ranging from A-F, we assigned the lower end of the letter grade range, for instance, A=93, A-=90, etc. Using an independent t-test, we also compared the mean grades for students who participated in the study against students in the course who did not. We obtained a mean of 82.55 for non-participant students and a mean of 83.55 for study participants, concluding that there was no significant difference (Table I). To validate the study population, we also ran a single sample t-test used for comparing the means of the study participants (single sample) against all the students enrolled in the course (population mean). Identical to the independent t-test findings, we found study participants reported statistically similar grades in the preceding course ($M=83.74$, $SD=5.48$) to the class population as a whole, ($t(50)=0.62$, $p=.54$).

Table I: Baseline comparison of course grades between study participants and nonparticipants.

Population	Measure	Mean	d.f.	Standard Error	t*exp*	p-value
Non-Participants	Previous Course	82.55	55	1.224	0.821	$.4<p<.5$
Participants		83.55	77	0.648		
Non-Participants	Current Course	82.96	45	1.087	1.355	$.1<p<.2$
Participants		84.64	59	0.782		

Similarly, we found no significant difference in the total course grades between non-participants and participants when comparing their grades in the current course (Table I). We ran an independent t-test of student grades in the previous course for those who continued to the current course. We obtained a mean of 82.96 for non-participants and 84.64 for participants (Table I). There were slight differences in the total sample size compared to the current course due to the handful of students who transferred in credit and did not take it at this institution. Additionally, for two students, final grades could not be procured. Again we fail to reject the null hypothesis (t-test expt=1.355, df=105), confirming our participants are representative of the population of students taking this course (Table I). We also used a single sample t-test to determine whether the mean course grade of the study participants was different from the class mean, a value determined by taking the average of the entire class. Study participants reported statistically similar grades in the course ($M=84.639$, $SD=6.1$) to the class population as a whole ($t(60)=.026$, $p=.36$).

In summary, we observed statistically similar scores for students' academic performance in the preceding course (Table I). In the current course, we observed a larger margin of difference than previously observed although still no significant difference between student performance in the two group types, (Table I).

APPENDIX G
Cooperative Lesson Planning Template
Modified from Johnson *et al*, 1998

Subject Area: Biology

Lesson: Modeling Changes in Atmospheric CO_2

Making Pre-instructional Decisions

Academic Objectives:

- Analyzing data and creating models of changes in atmospheric CO_2 using Microsoft Excel as a tool.
- Generating quality figures and writing appropriate figure captions.

Social Skills Objectives:

- *Integrate Ideas Into a Single Position:* The lesson is structured so group members carry out an Excel modeling activity together, and then synthesize their ideas on how best to summarize their generated data in a figure and an appropriate figure caption. The complete models (Figure + Figure captions) must be completed in groups of 3 and submitted as a group assignment.

Group Size: 3 Students per group.

Method of Assigning Students: Before the beginning of the unit, students of varying Excel experience will be paired together determined from a pre-test/pre-activity.

Roles:

1) *Guide and Note Taker:* The student closely monitors the procedural steps
2) *Excel Navigator:* The student navigates the Microsoft Excel tool to put into action the procedural steps.
3) *Checker/Explainer*: Synthesize the group's interpretation of the data generated and create a written record of all the group's responses to a shared Google document.

Room Arrangement: Students work in small groups in different areas around the classroom.

Materials: Materials for activity. Homework question sheet, laptop/table for figure generation.

- Computer with Microsoft Office Suite Package
- One copy of the Activity

1. Explaining Task And Cooperative Goal Structure

Students will generate two models that incorporate various sources of empirically measured data of CO_2 concentrations, the quantity of CO_2 being *emitted* (added) into the atmosphere, and the quantity that is being removed by vegetation and the oceans. These models will be constructed using Excel as a tool. Upon completion, students submit one copy of the completed models along with a summary of each model to the teacher. (Williams & Wassenberg, 2017)

2. Criteria For Success:

- Before the start of the activity, the students are informed by the instructor that they will receive a collective grade for the group.
- The activity is time-based, encouraging students to coordinate and communicate effectively with one another.

3. Positive Interdependence:

Positive Task Interdependence: The activity requires both the guide, explainer and the navigator to exercise a division of labor. The guide has to complete their action before the navigator or the explainer can complete their responsibility.
Goal Interdependence: Students have to be interdependent to accomplish the goal of completing the activity. The resulting assignment is set up so that students can only do it if all members attain their individual goals. The assignment is also submitted as a group to encourage goal interdependence.
Positive Resource Interdependence: Only a single copy of the assignment is handed to each group encouraging members to share and complete the assignment together.
Positive Role Interdependence: Roles are clearly defined in the activity introduction and prompt for the lesson.

4. Individual Accountability:

- The instructor will gauge each student's understanding by randomly asking questions related to the assignment (both conceptual and procedural) to members of the group throughout the activity.
- At the end of the activity, the teacher will set up a short post quiz on the topic to gauge each member’s understanding. This post quiz will be compared to the

pre-quiz completed by each member to ensure that individual learning is in fact benefitted by working in a cooperative group.

5. **Intergroup Cooperation:**

   - **Striving for a mutual benefit:** Once each group has completed their assignment, the teacher will implement a structured peer review where groups will have an opportunity to provide written feedback on the models of another group. The peer review of the group work will be submitted alongside the assignment. If done well, groups will receive credit for completing the peer review in addition to credit for completing their assignment.

6. **Expected Behaviors:**

   - Group members will be encouraged to listen to each other, stay on task, and complete the assignment. Members should be providing positive feedback to each other and when appropriate providing constructive feedback, both beneficial to a cooperative learning environment.

Monitoring And Intervening

1. **Observation Procedure:** Informally

2. **Observation By:** Teacher

3. **Intervening for Task Assistance:**

   - The teacher walks around the room to check on each group, answer clarifying questions or help students along with the assignment.
   - When needed, the teacher uses scaffolding techniques such as prompts or questions to gauge student understanding.
   - To test understanding, the instructor can verbally ask a problem-solving question such as, “ What do you think might be the causes for the differences in the measured versus the calculated values?” Being able to demonstrate problem-solving skills demonstrates mastery of the topic.

4. **Intervening for Teamwork Assistance:**

   - The teacher is constantly walking around to teach group monitoring so that all group members are carrying forth the task at hand. If a teacher notes a number of minor issues of students' expected behaviors, the teacher will make sure to get all the group's attention and make a general announcement about the classroom group expectations.

- If group members continue to sway away from the expectations, the teacher will intervene with that group specifically. The teacher will work with the group to identify the issue at hand and find ways to support the learning of the members.

5. **Other:** ____

Assessing and Processing

1. **Assessment of Members' Individual Learning:**

   - Individual Quiz (end of class) and random call-out questions from the teacher throughout the lesson.

2. **Assessment of Group Productivity:**

   - The teacher informally monitors group progress.
   - Groups complete activity in the given time.

3. **Small Group Processing:**

   - To each group member at the end of the lesson:
     - One thing you did that helped me understand modeling was...
     - One thing you did that helped me understand the influences on the calculation of atmospheric CO_2 is...

4. **Whole-Class Processing:**

   - One thing you learned from the presenting group that helped you in your understanding
   - One thing you learned from the presenting group that you will implement in your own future presentations.

5. **Charts and Graphs Used:**

   - The teacher has templates of different data sets to show examples of high-quality figures generated in excel and examples that need improving. Students can complete this short exercise before beginning their activity to know what constitutes a high-quality figure.

6. **Positive Feedback to Each Student:**

   - At the end of the lesson or unit, each member gives the other written positive feedback.

- The teacher can provide positive feedback to the whole class on areas where they observed good use of cooperative skills with examples from the cooperative groups.

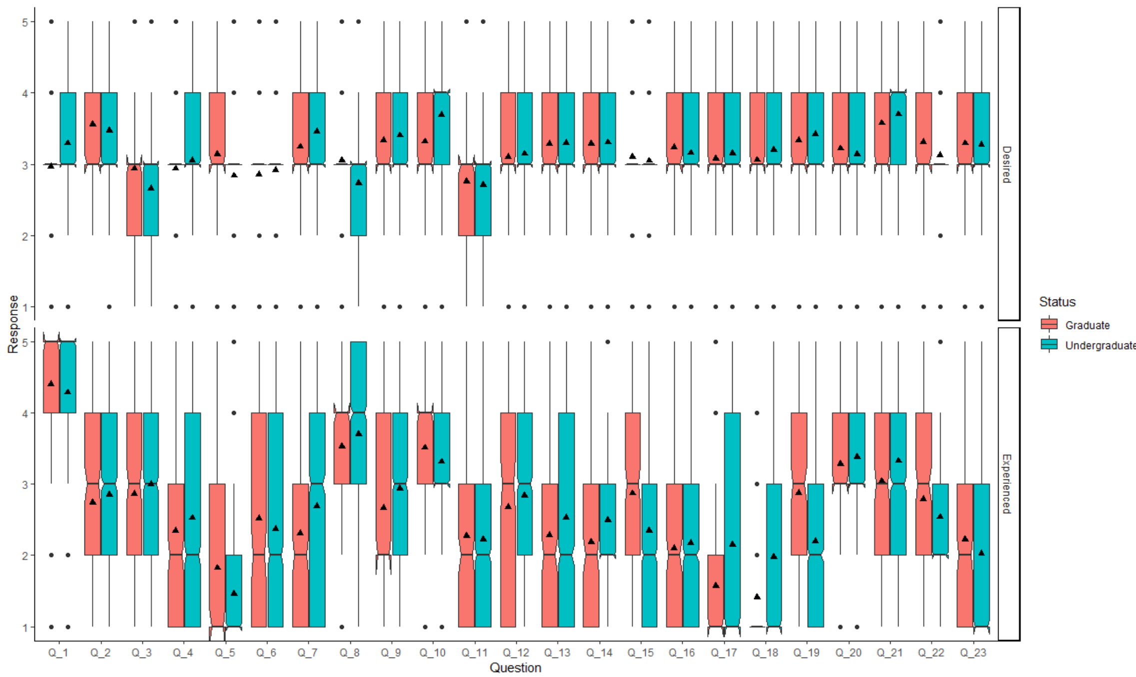
7. **Goal Setting for Improvement:**

- At the end of the lesson, members provide one constructive feedback to group members with a possible strategy on how they can improve.
- Students then process the feedback. Before beginning a new lesson, group members meet up at least once to arrive on the same page on how to implement the change.

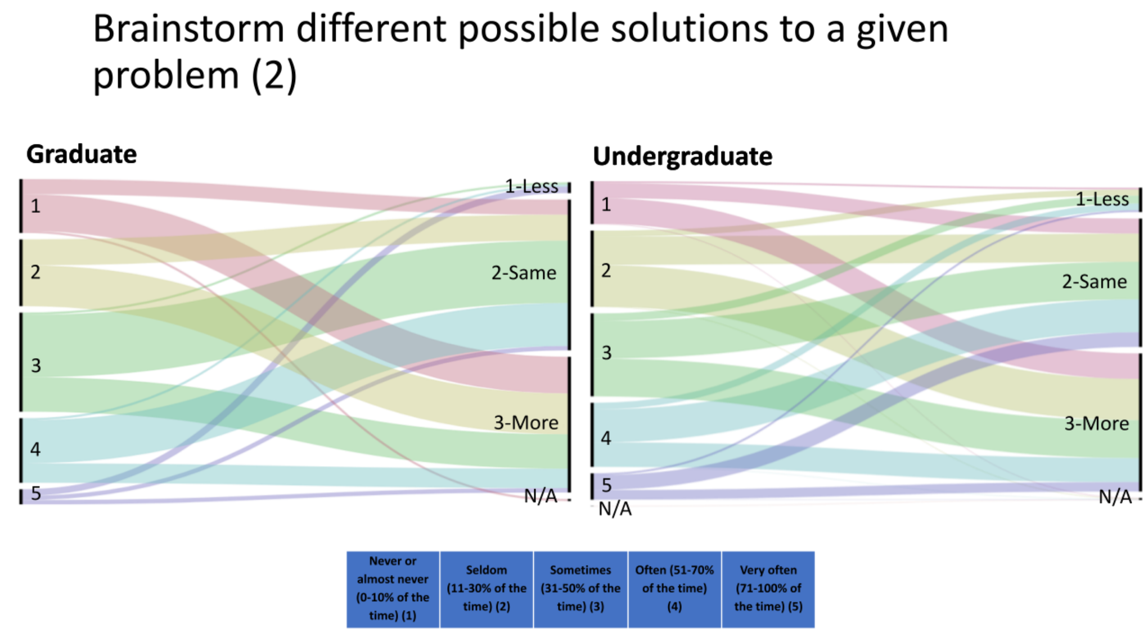
8. **Celebration:** If all students complete all the tasks (activity and a peer review), the teacher gives students verbal praise for their work and credit for completing the peer review also.

9. **Other:** ____

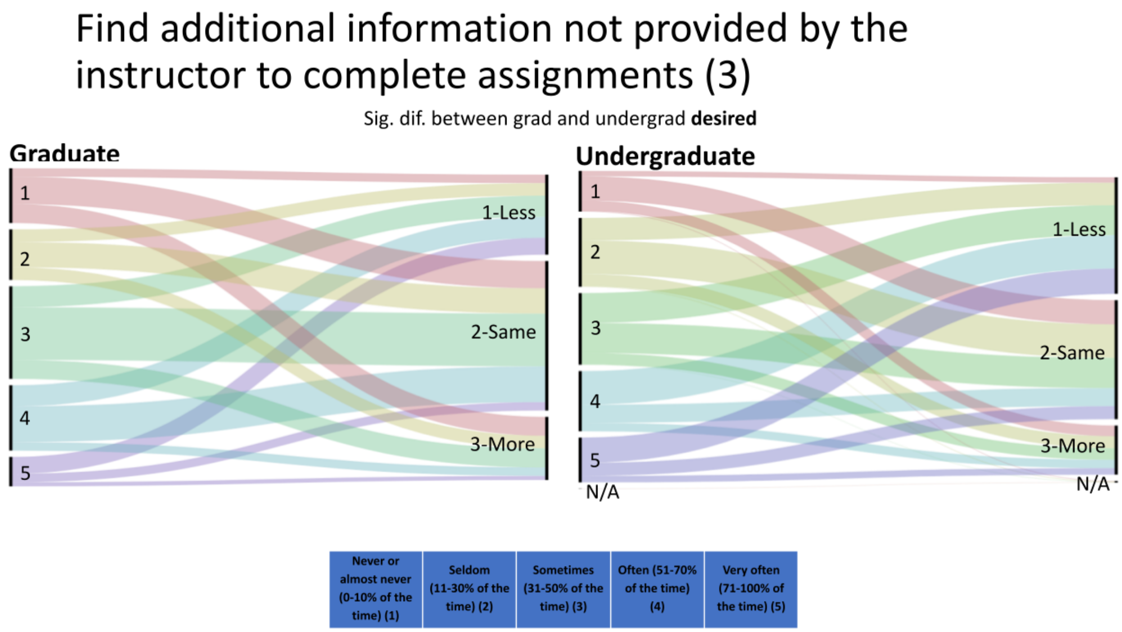
SUPPLEMENTARY FIGURES



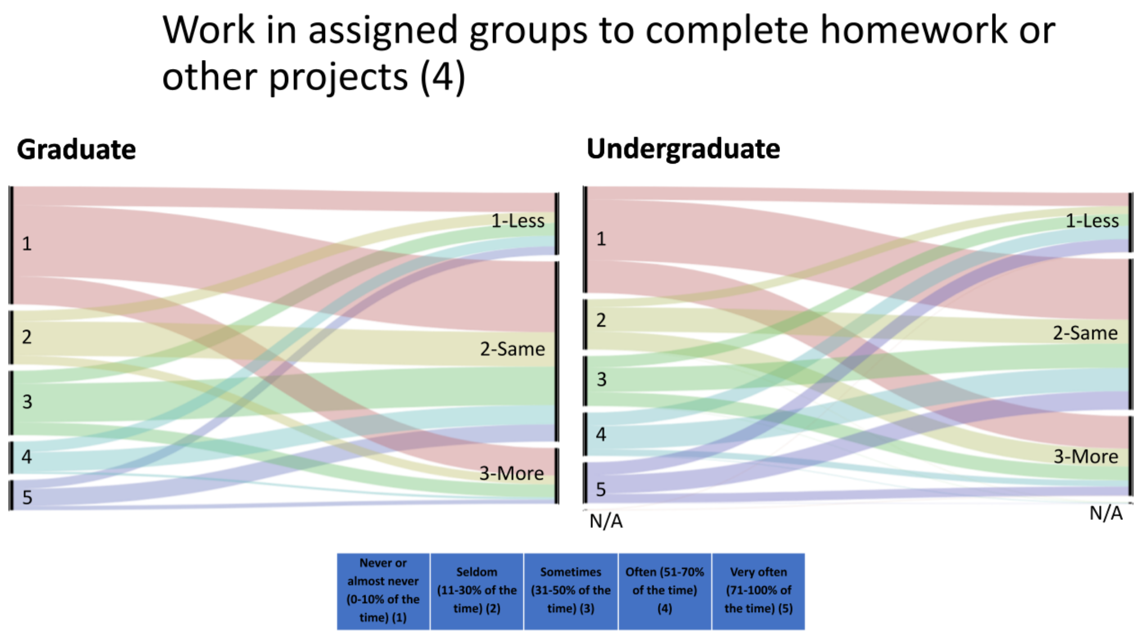
S1 Fig: Box and whisker plots of all surveyed instructional practices (Q_1 through Q_23) reported by graduate (*n*=161) and undergraduate (*n*=1113) students in STEM classrooms. Upper row (Desired) details levels of desire for each instructional activity. Lower row (Experienced) details levels of each instructional activity reported as actually experienced in the classroom. Levels range from 1 (Never or almost never; 0-10% of the time) to 5 (Very often; 71-100% of the time). Triangles indicate mean values.



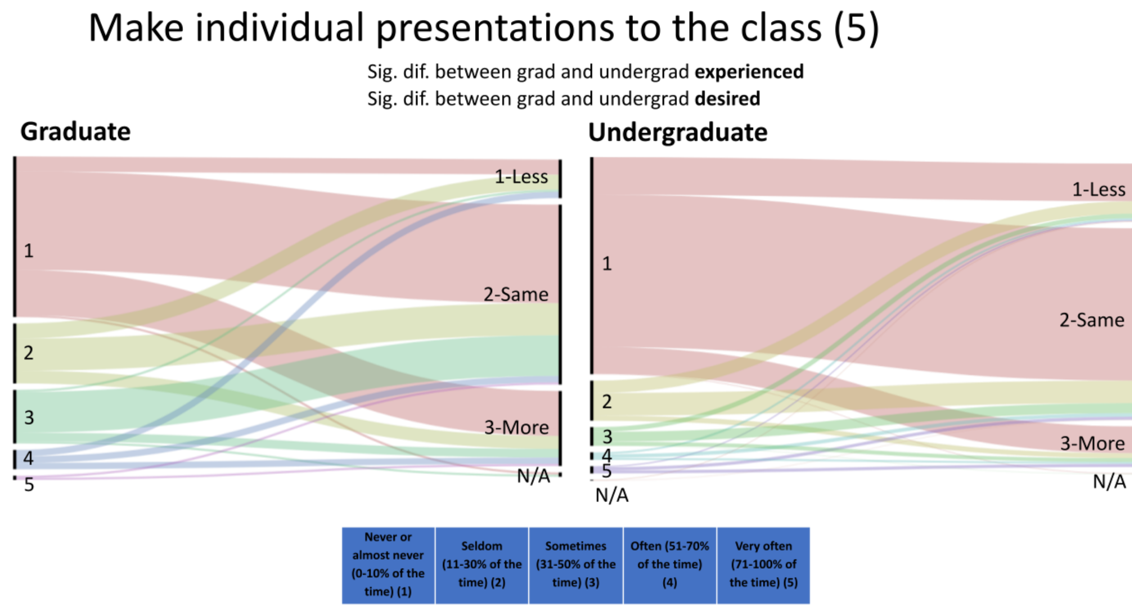
S2 Fig: Alluvial plots of how often students reported experiencing the teaching practice in Q_2 ("*Brainstorm different possible solutions to a given problem*") and how much that teaching practice was desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left side of each graph) and; the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia" or lines between columns indicate how often students reported experiencing the teaching practice (left side of each graph) and how much this teaching practice was desired (right side of each graph) and are proportional to the number of respondents. NA= respondents who did not respond to a particular question. 1=Never or almost never (0-10% of the time); 2=Seldom (11-30% of the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time).



S3 Fig: Alluvial plots of how often students reported experiencing the teaching practice in Q_3 ("*Find additional information not provided by the instructor to complete assignments*") and how much that teaching practice was desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left side of each graph) and; the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia" or lines between columns indicate how often students reported experiencing the teaching practice (left side of each graph) and how much this teaching practice was desired (right side of each graph) and are proportional to number of respondents. NA= respondents who did not respond to a particular question. 1=Never or almost never (0-10% of the time); 2=Seldom (11-30% of the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time).

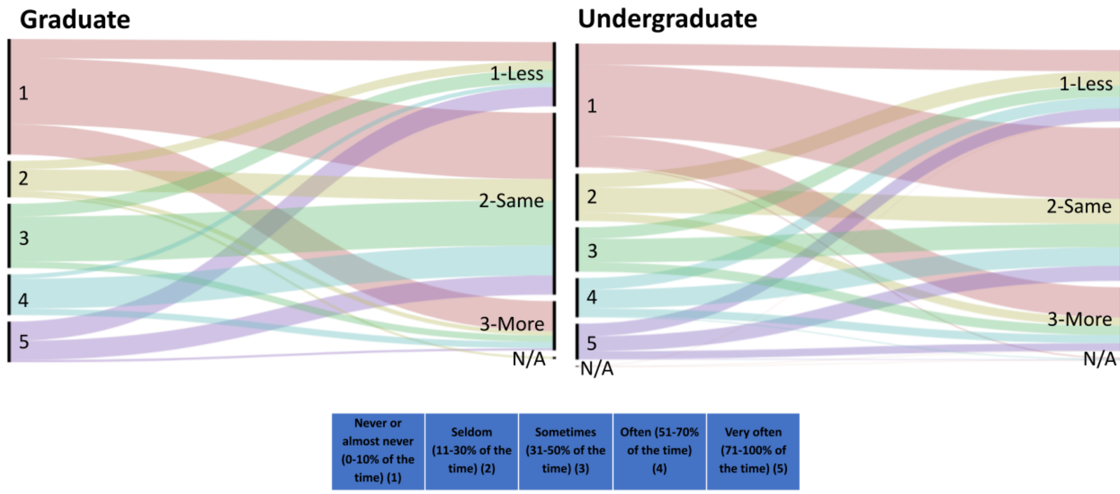


S4 Fig: Alluvial plots of how often students reported experiencing the teaching practice in Q_4 ("*Work in assigned groups to complete homework or other projects*") and how much that teaching practice was desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left side of each graph) and; the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia" or lines between columns indicate how often students reported experiencing the teaching practice (left side of each graph) and how much this teaching practice was desired (right side of each graph) and are proportional to the number of respondents. NA= respondents who did not respond to a particular question. 1=Never or almost never (0-10% of the time); 2=Seldom (11-30% of the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time).

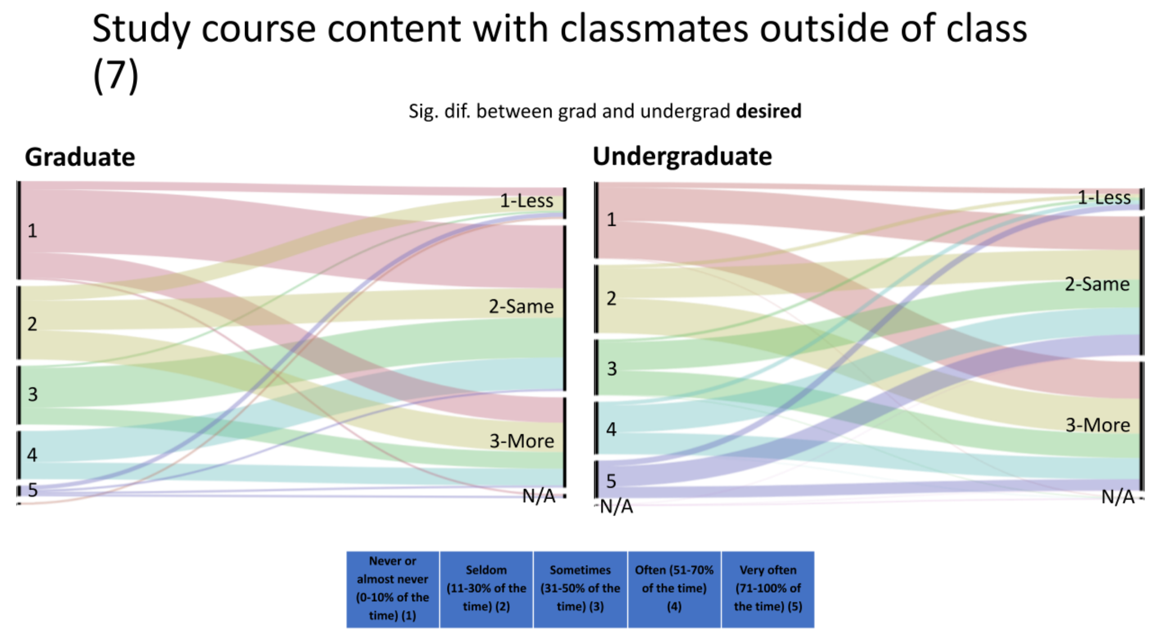


S5 Fig: Alluvial plots of how often students reported experiencing the teaching practice in Q_5 ("*Make individual presentations to the class*") and how much that teaching practice was desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left side of each graph) and; the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia" or lines between columns indicate how often students reported experiencing the teaching practice (left side of each graph) and how much this teaching practice was desired (right side of each graph) and are proportional to the number of respondents. NA= respondents who did not respond to a particular question. 1=Never or almost never (0-10% of the time); 2=Seldom (11-30% of the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time).

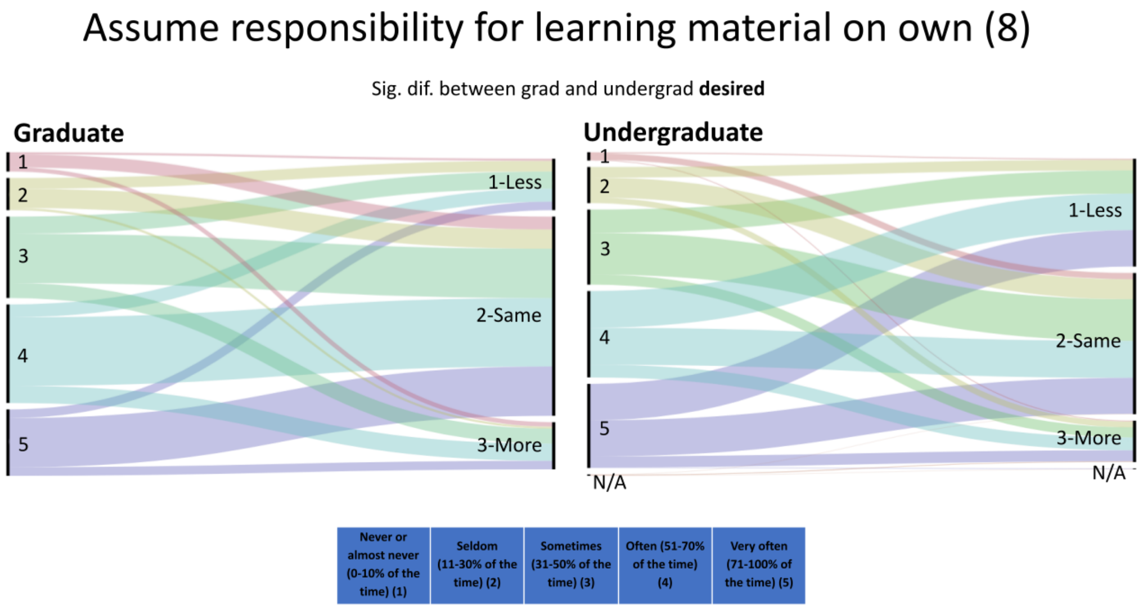
Be graded on class participation (6)



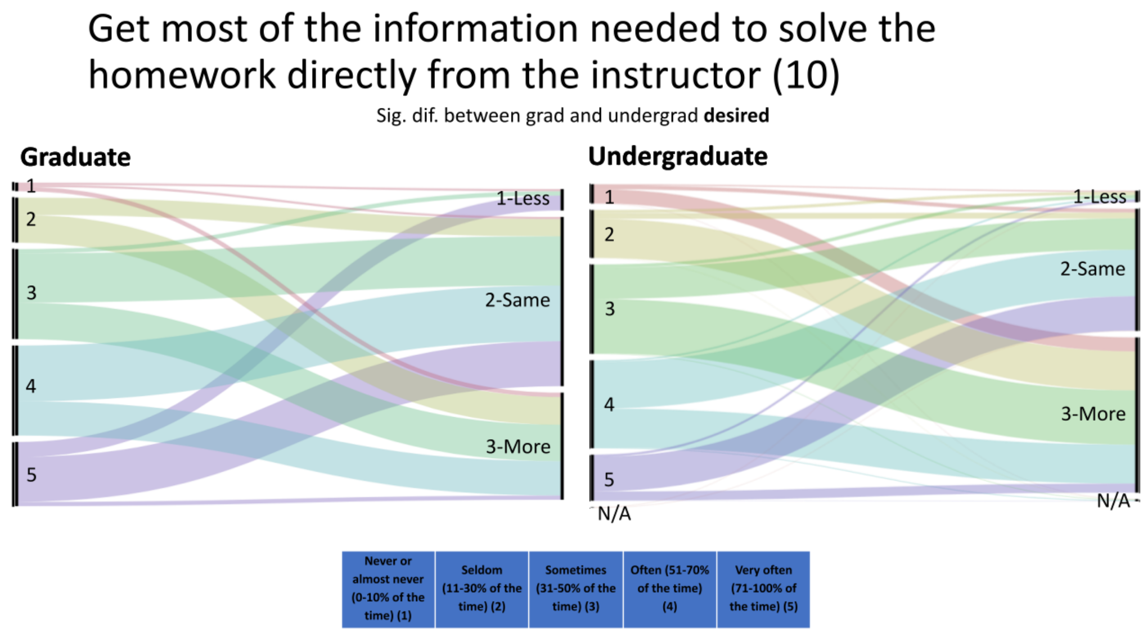
S6 Fig: Alluvial plots of how often students reported experiencing the teaching practice in Q_6 ("*Be graded on class participation*") and how much that teaching practice was desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left side of each graph) and; the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia" or lines between columns indicate how often students reported experiencing the teaching practice (left side of each graph) and how much this teaching practice was desired (right side of each graph) and are proportional to the number of respondents. NA= respondents who did not respond to a particular question. 1=Never or almost never (0-10% of the time); 2=Seldom (11-30% of the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time).



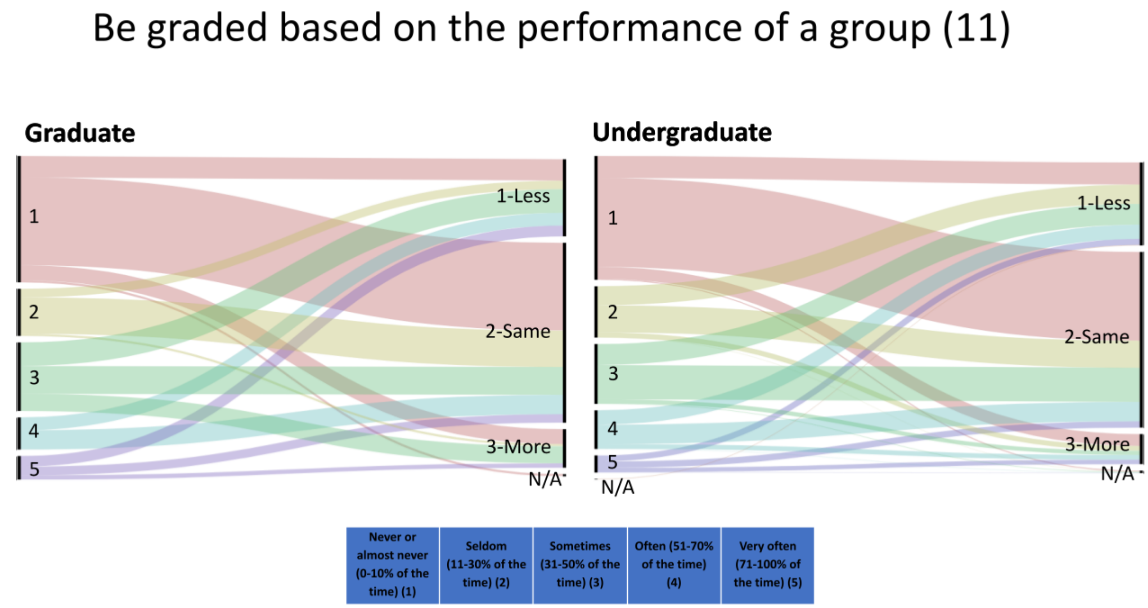
S7 Fig: Alluvial plots of how often students reported experiencing the teaching practice in Q_7 ("*Study course content with classmates outside of class*") and how much that teaching practice was desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left side of each graph) and; the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia" or lines between columns indicate how often students reported experiencing the teaching practice (left side of each graph) and how much this teaching practice was desired (right side of each graph) and are proportional to the number of respondents. NA= respondents who did not respond to a particular question. 1=Never or almost never (0-10% of the time); 2=Seldom (11-30% of the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time).



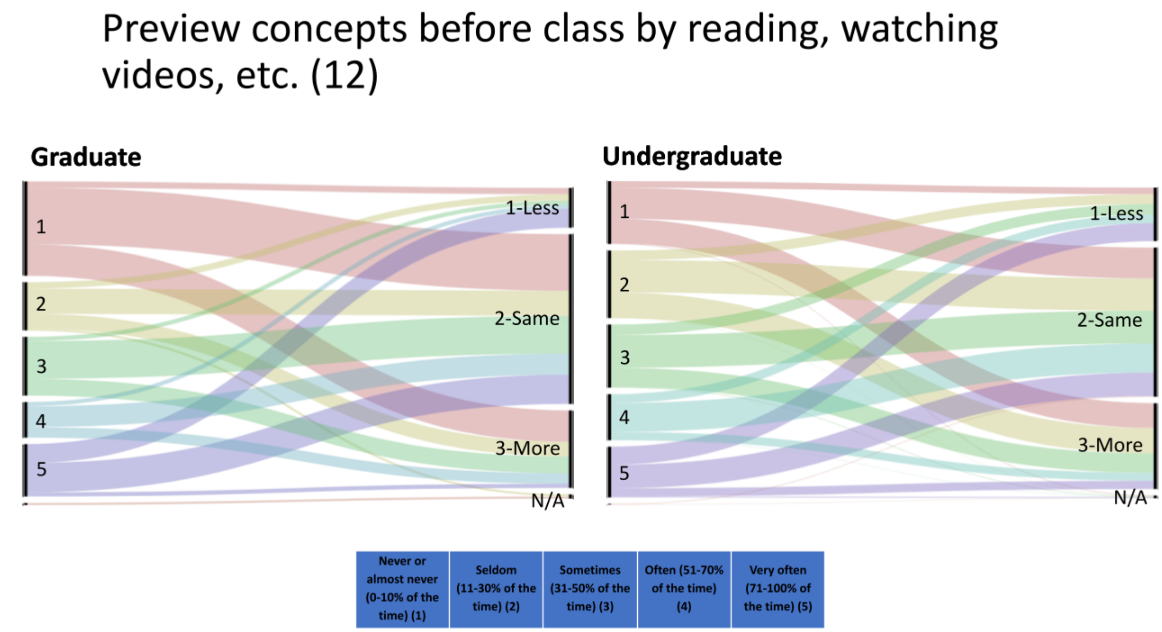
S8 Fig: Alluvial plots of how often students reported experiencing the teaching practice in Q_8 ("*Assume responsibility for learning material on own*") and how much that teaching practice was desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left side of each graph) and; the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia" or lines between columns indicate how often students reported experiencing the teaching practice (left side of each graph) and how much this teaching practice was desired (right side of each graph) and are proportional to the number of respondents. NA= respondents who did not respond to a particular question. 1=Never or almost never (0-10% of the time); 2=Seldom (11-30% of the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time).



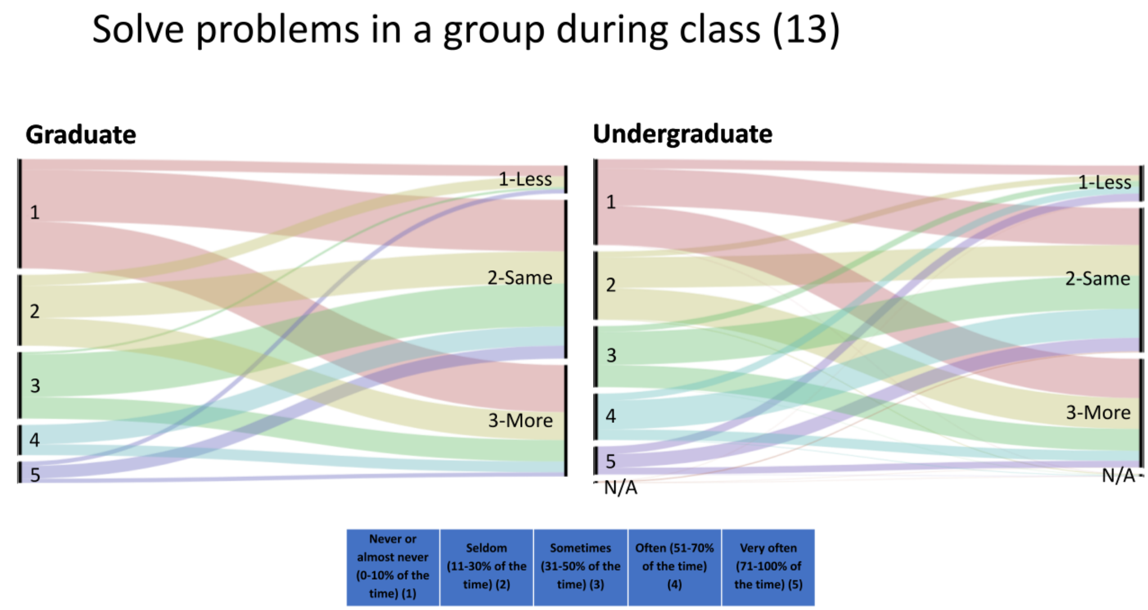
S9 Fig: Alluvial plots of how often students reported experiencing the teaching practice in Q_10 ("*Get most of the information needed to solve the homework directly from the instructor*") and how much that teaching practice was desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left side of each graph) and; the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia" or lines between columns indicate how often students reported experiencing the teaching practice (left side of each graph) and how much this teaching practice was desired (right side of each graph) and are proportional to the number of respondents. NA= respondents who did not respond to a particular question. 1=Never or almost never (0-10% of the time); 2=Seldom (11-30% of the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time).



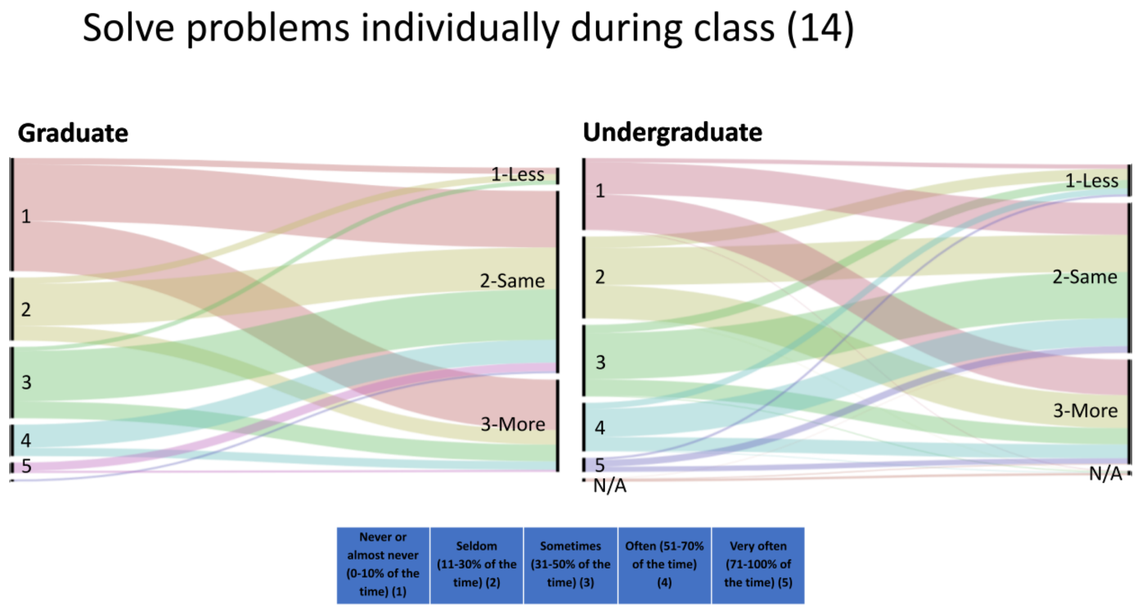
S10 Fig: Alluvial plots of how often students reported experiencing the teaching practice in Q_11 ("*Be graded based on the performance of a group*") and how much that teaching practice was desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left side of each graph) and; the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia" or lines between columns indicate how often students reported experiencing the teaching practice (left side of each graph) and how much this teaching practice was desired (right side of each graph) and are proportional to the number of respondents. NA= respondents who did not respond to a particular question. 1=Never or almost never (0-10% of the time); 2=Seldom (11-30% of the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time).



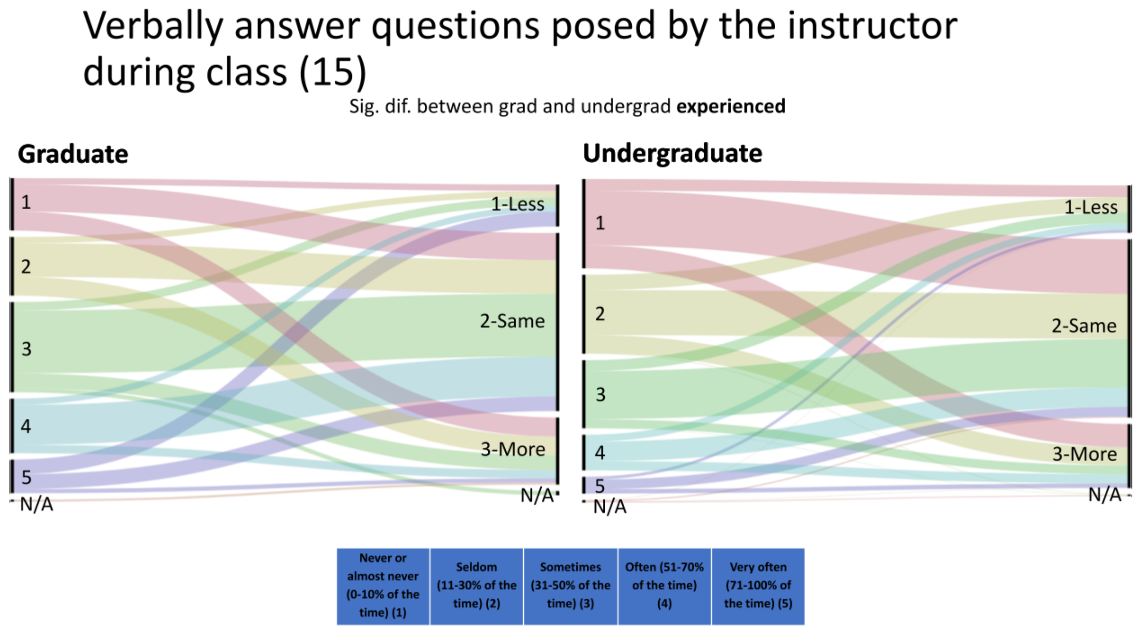
S11 Fig: Alluvial plots of how often students reported experiencing the teaching practice in Q_12 ("*Preview concepts before class by reading, watching videos, etc.*") and how much that teaching practice was desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left side of each graph) and; the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia" or lines between columns indicate how often students reported experiencing the teaching practice (left side of each graph) and how much this teaching practice was desired (right side of each graph) and are proportional to the number of respondents. NA= respondents who did not respond to a particular question. 1=Never or almost never (0-10% of the time); 2=Seldom (11-30% of the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time).



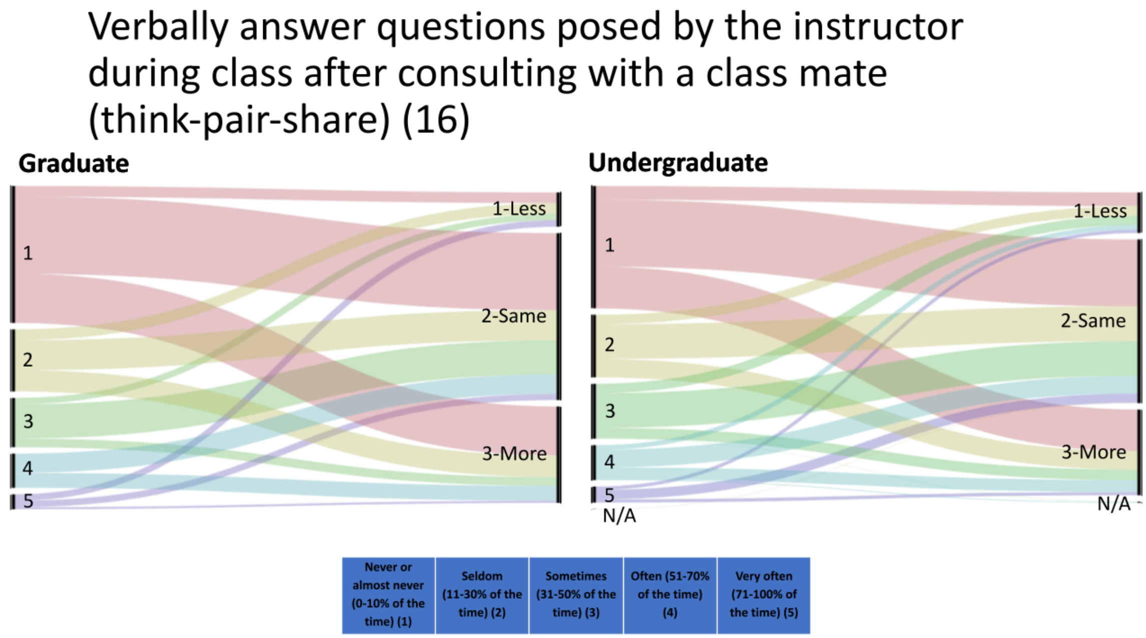
S12 Fig: Alluvial plots of how often students reported experiencing the teaching practice in Q_13 ("*Solve problems in a group during class*") and how much that teaching practice was desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left side of each graph) and; the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia" or lines between columns indicate how often students reported experiencing the teaching practice (left side of each graph) and how much this teaching practice was desired (right side of each graph) and are proportional to the number of respondents. NA= respondents who did not respond to a particular question. 1=Never or almost never (0-10% of the time); 2=Seldom (11-30% of the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time).



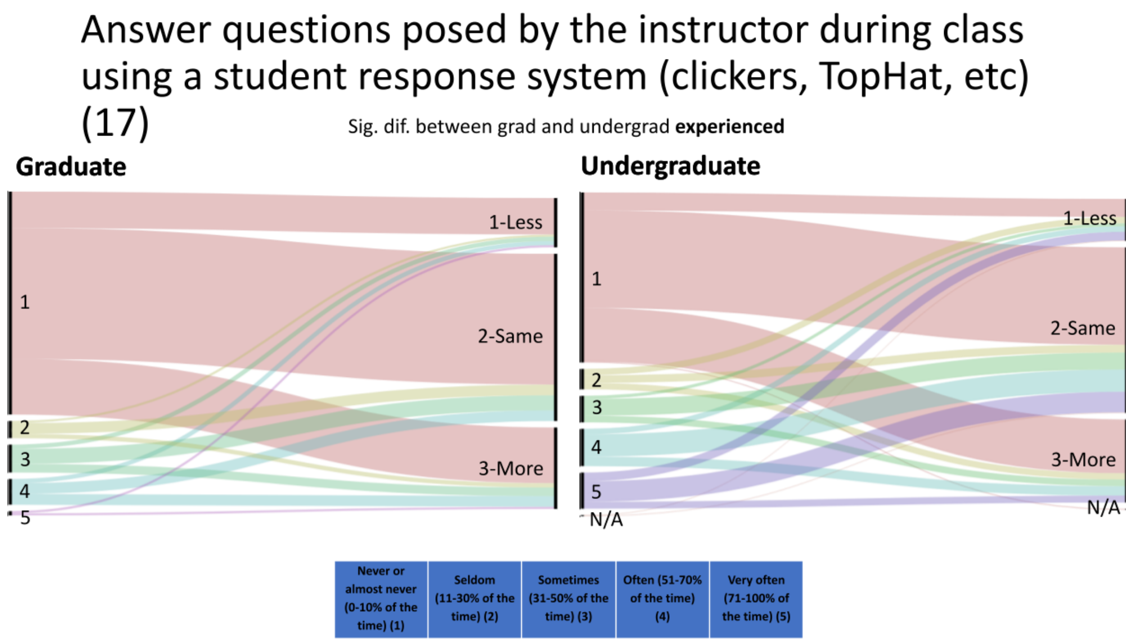
S13 Fig: Alluvial plots of how often students reported experiencing the teaching practice in Q_14 ("*Solve problems individually during class*") and how much that teaching practice was desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left side of each graph) and; the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia" or lines between columns indicate how often students reported experiencing the teaching practice (left side of each graph) and how much this teaching practice was desired (right side of each graph) and are proportional to the number of respondents. NA= respondents who did not respond to a particular question. 1=Never or almost never (0-10% of the time); 2=Seldom (11-30% of the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time).



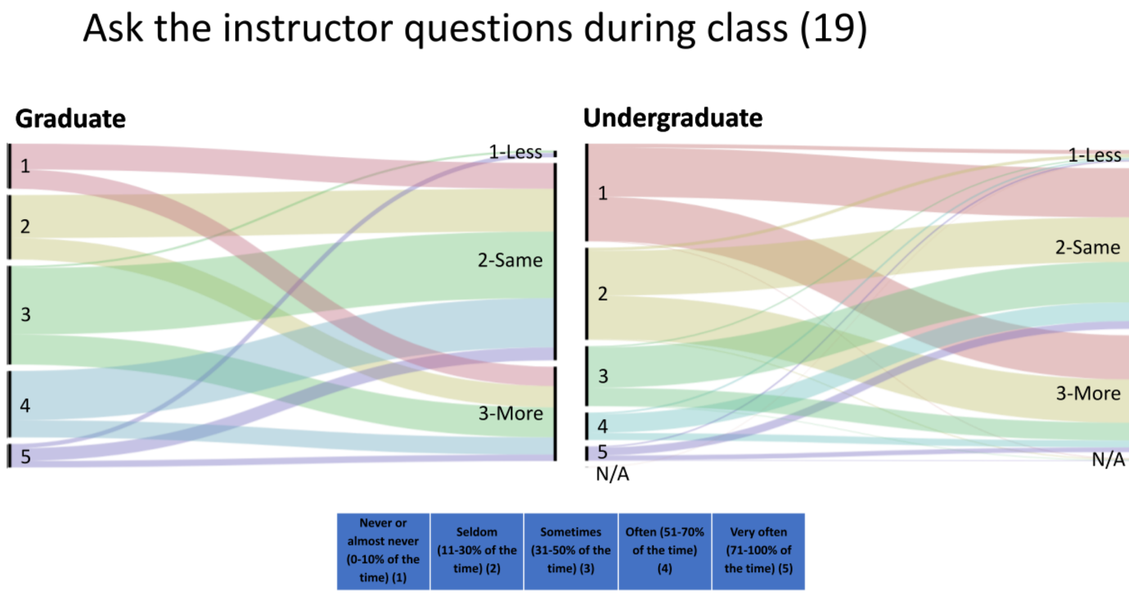
S14 Fig: Alluvial plots of how often students reported experiencing the teaching practice in Q_15 ("*Verbally answer questions posed by the instructor during class*") and how much that teaching practice was desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left side of each graph) and; the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia" or lines between columns indicate how often students reported experiencing the teaching practice (left side of each graph) and how much this teaching practice was desired (right side of each graph) and are proportional to the number of respondents. NA= respondents who did not respond to a particular question. 1=Never or almost never (0-10% of the time); 2=Seldom (11-30% of the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time).



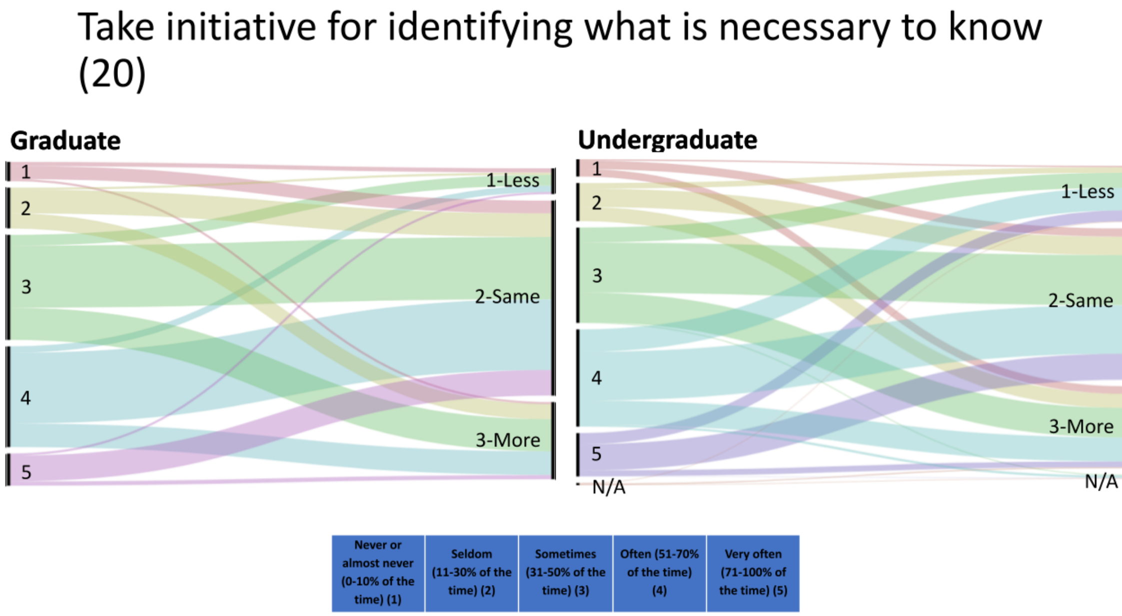
S15 Fig: Alluvial plots of how often students reported experiencing the teaching practice in Q_16 ("*Verbally answer questions posed by the instructor during class after consulting with a classmate (think-pair-share)*") and how much that teaching practice was desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left side of each graph) and; the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia " or lines between columns indicate how often students reported experiencing the teaching practice (left side of each graph) and how much this teaching practice was desired (right side of each graph) and are proportional to the number of respondents. NA= respondents who did not respond to a particular question. 1=Never or almost never (0-10% of the time) ; 2=Seldom (11-30% of the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time).



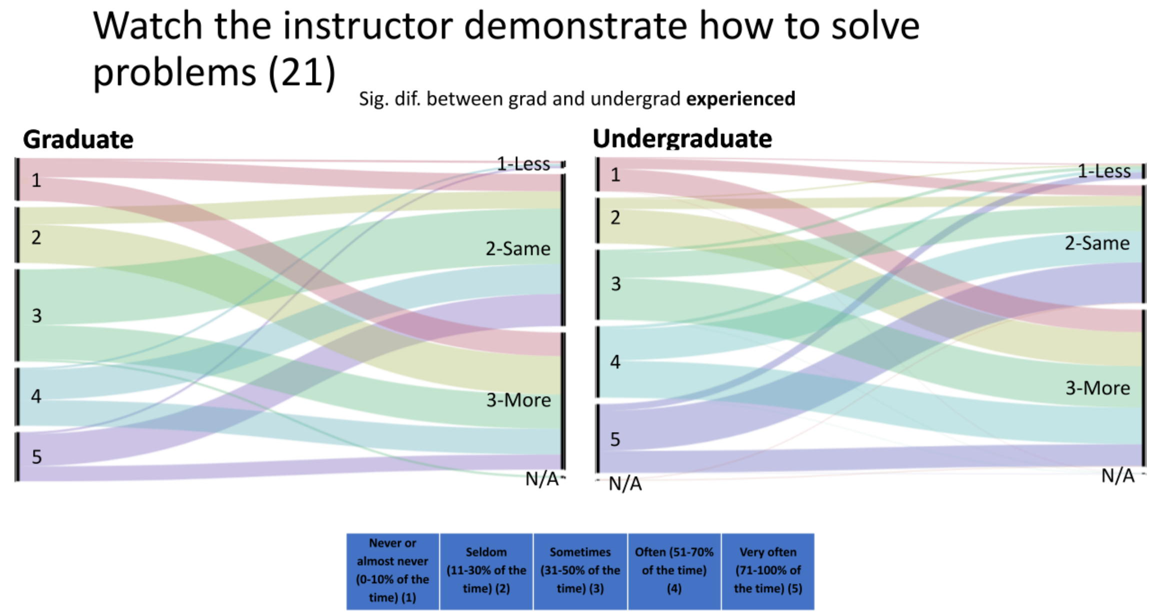
S16 Fig: Alluvial plots of how often students reported experiencing the teaching practice in Q_17 ("*Answer questions posed by the instructor during class using a student response system (clickers, TopHat, etc)*") and how much that teaching practice was desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left side of each graph) and; the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia" or lines between columns indicate how often students reported experiencing the teaching practice (left side of each graph) and how much this teaching practice was desired (right side of each graph) and are proportional to the number of respondents. NA= respondents who did not respond to a particular question. 1=Never or almost never (0-10% of the time) ; 2=Seldom (11-30% of the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time).



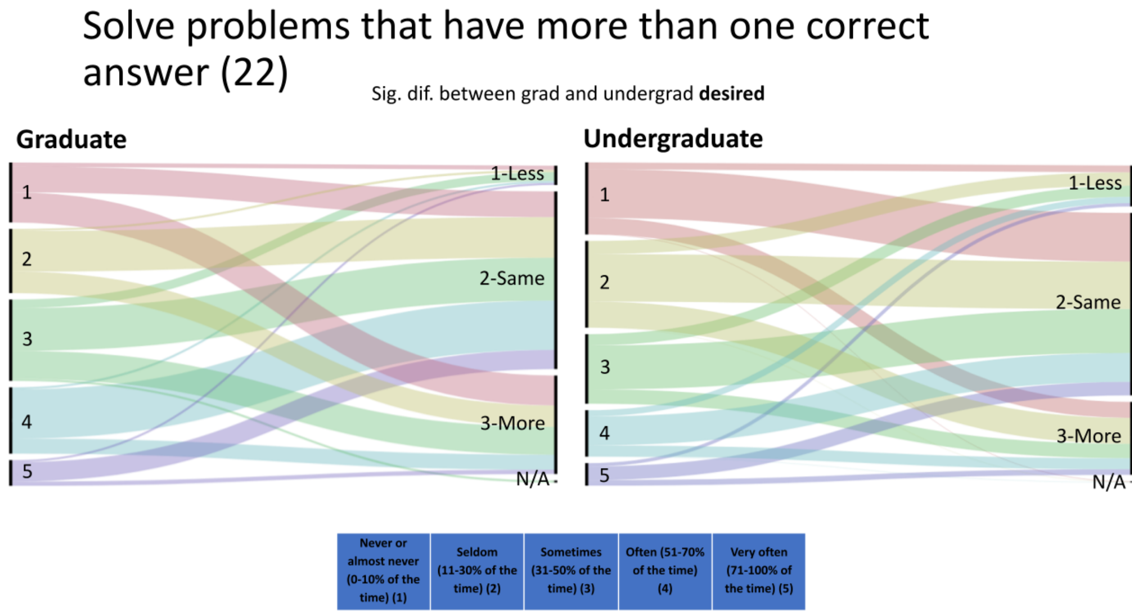
S17 Fig: Alluvial plots of how often students reported experiencing the teaching practice in Q_19 ("*Ask the instructor questions during class*") and how much that teaching practice was desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left side of each graph) and; the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia" or lines between columns indicate how often students reported experiencing the teaching practice (left side of each graph) and how much this teaching practice was desired (right side of each graph) and are proportional to the number of respondents. NA= respondents who did not respond to a particular question. 1=Never or almost never (0-10% of the time); 2=Seldom (11-30% of the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time).



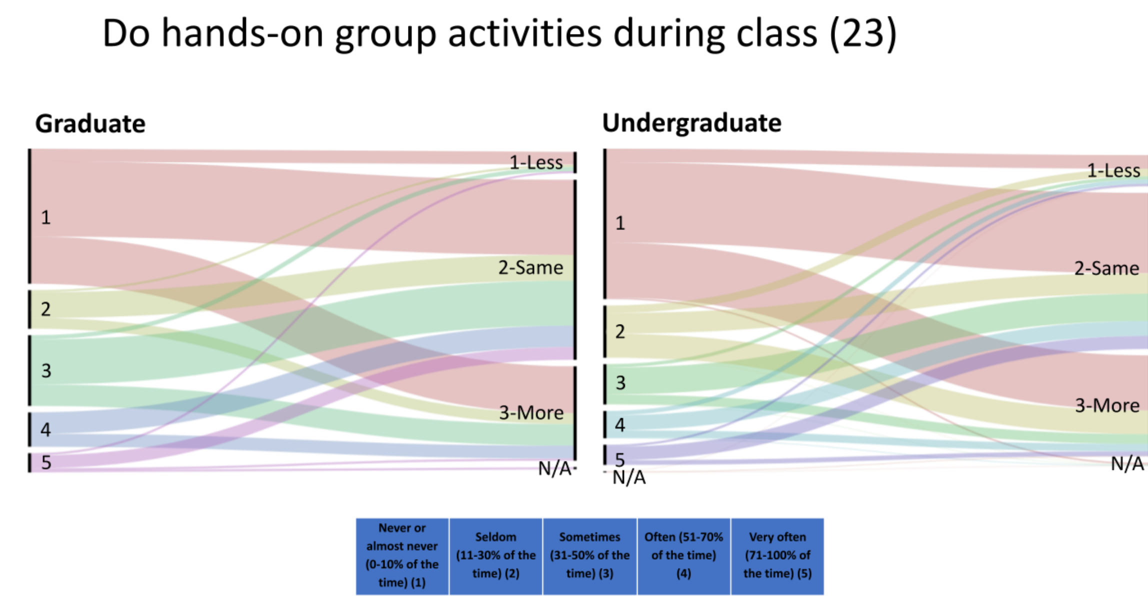
S18 Fig: Alluvial plots of how often students reported experiencing the teaching practice in Q_20 ("*Take initiative for identifying what is necessary to know*") and how much that teaching practice was desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left side of each graph) and; the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia" or lines between columns indicate how often students reported experiencing the teaching practice (left side of each graph) and how much this teaching practice was desired (right side of each graph) and are proportional to the number of respondents. NA= respondents who did not respond to a particular question. 1=Never or almost never (0-10% of the time); 2=Seldom (11-30% of the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time).



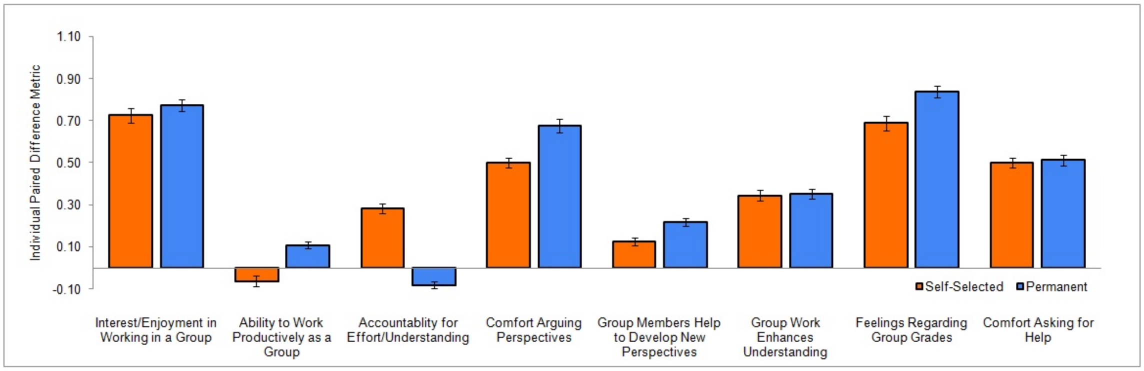
S19 Fig: Alluvial plots of how often students reported experiencing the teaching practice in Q_21 ("*Watch the instructor demonstrate how to solve problems*") and how much that teaching practice was desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left side of each graph) and; the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia" or lines between columns indicate how often students reported experiencing the teaching practice (left side of each graph) and how much this teaching practice was desired (right side of each graph) and are proportional to the number of respondents. NA= respondents who did not respond to a particular question. 1=Never or almost never (0-10% of the time); 2=Seldom (11-30% of the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time).



S20 Fig: Alluvial plots of how often students reported experiencing the teaching practice in Q_22 ("*Solve problems that have more than one correct answer*") and how much that teaching practice was desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left side of each graph) and; the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia" or lines between columns indicate how often students reported experiencing the teaching practice (left side of each graph) and how much this teaching practice was desired (right side of each graph) and are proportional to the number of respondents. NA= respondents who did not respond to a particular question. 1=Never or almost never (0-10% of the time); 2=Seldom (11-30% of the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time).



S21 Fig: Alluvial plots of how often students reported experiencing the teaching practice in Q_23 ("*Do hands-on group activities during class*") and how much that teaching practice was desired for graduate students (left column) and undergraduates (right column). The vertical bars are the experienced frequency (left side of each graph) and; the vertical height of each bar is proportional to the number of respondents who chose that response. The "alluvia" or lines between columns indicate how often students reported experiencing the teaching practice (left side of each graph) and how much this teaching practice was desired (right side of each graph) and are proportional to the number of respondents. NA= respondents who did not respond to a particular question. 1=Never or almost never (0-10% of the time); 2=Seldom (11-30% of the time); 3=Sometimes (31-50% of the time); 4=Often (51-70% of the time); 5=Very often (71-100% of the time).



S22 Fig. Top five most frequent codes with the greatest change (Individual Paired Difference) between pre and post semester responses for Self-Selected (Orange) and assigned Groups (Blue). The top codes for self-selected include: 1) Comfort Asking for Help 2) Interest/Enjoyment in Working in a Group 3) Feelings Regarding Group Grades 4) Enhanced Understanding (4.22 to 4.56- ss), (4.19 to 4.54-p, rank 5) Top codes for assigned include 1) Feelings Regarding Group Grades 2) Interest/Enjoyment in Working in a Group 3) Comfort Arguing Perspectives 4) Comfort asking for help.

S Table 1. Categories and subcategories identified in student responses to the prompt "*Please explain your overall perceptions of the utility of group work.*"

Categories	Sub-Categories
Bonding Experience	Safe space Enjoyed group work
Collaboration Skills	Collaboration Participating equally Helpful/Useful/Beneficial Accountability Efficiency
Understanding Material	Content understanding Diverse perspectives Motivation Learning from each other New ways of understanding
Negative Experience	Stressful/Frustrating Jeopardizes grades Beneficial with equal participation Preferred individual work
Personal Growth	Learning to work in a group Life skills Communicational skills Building confidence Practice maturity

S Table 2. Categories (Benefits and Challenges) identified in student post semester responses to the prompt "*From your perspective, what were the benefits and challenges associated with your group*?"

Group Type	Benefits	No Benefits	Challenges	No Challenge	Sum
Assigned	51	0	23	15	89
Self-Selected	35	0	20	9	64

Note. Summed total of identified categories in assigned and self-selected groups from the Post- Semester Response.

S Table 3. Categories and subcategories identified from Pre and Post-Semester student responses to the prompt "*Please specify any concerns you have about continuing to work in a group.*"

Categories	Sub-Categories
Unequal Work/Participation	*People don't participate* *Unequal Contribution* *Accountability* *Doing all/ majority of the work*
Mindset	*work ethic* *people feel less responsible*
Interest Incongruency	*Motivation/ determination/ engaged/ dedicated* *People don't care/ low interest/ apathetic*
Effect on Course Grades	*Grade based off group work - potentially harmful/ grades suffer* *group evaluations (should affect grades)* *refuse obligations*
Hinder Learning	*difficult to understand content in group setting* *left behind*
Group Dynamics	*uncomfortable/ don't get along* *trust issues (others not doing work correctly)* *counterproductive/back-tracking/ holding back others* *refuse obligations* *choose groups* *underprepared*

S Table 4. Categories identified in student elaboration responses to the prompt "*How do you feel about having some of the assignments based on group performance?"*

	Categories			
Group Type	Positive	Negative	Constructive	Neutral
Assigned	8	20	8	4
Self-Selected	9	12	10	2

Note. Summed total of identified categories in assigned and self-selected groups in Pre-Semester Response.

S Table 5. Categories and subcategories identified in student elaboration responses to the prompt "*Working in my group has helped in my understanding of the laboratory material."*

A.	**Categories**		
Group Type	Agree	Disagree	Neither
Self-Selected	19	2	1
Assigned	26	5	3

B.	**Categories**		
Group Type	Agree	Disagree	Neither
Self-Selected	16	1	1
Assigned	24	1	2

C.		**Self-Selected**		**Assigned**	
Categories	Subcategories	Pre	Post	Pre	Pre
AGREE	Group members enhanced learning and understanding of the subject	*"Having people explain how they think of problems and solutions helps me better understand questions. It also gives me a larger tool box to work out of"*	*"There were many times when I had a group member explain a different point of view to me where the point of view helped me understand the material in a different way"*	*"My partner(s) (previous course) helped clarify things I was confused about, and made it more motivating to work hard because they were also."*	*"Yes, I understood concepts a lot better with group work because if I didn't understand one perspective, another person would try to explain it to me"*
	Contribution of diverse ideas about material	*"Group members bring other perspectives and insight that I didnâ€™t have myself." 5*	*"I learn a lot more effectively when I'm actively talking about the material and not just bobble-heading like 'yeah, sure, uh huh, I know what's going on.'" 10*	*"They could sometimes explain things in a way that the professor could not."*	*"I felt like their understandings helped me to solidify my own because we all had different perspectives and could piece together what we did and didn't know to figure out the answers to a question."*

Note. Summed total of identified categories in assigned and self-selected groups A) Pre-Semester Response B) Post-Semester Response C) Sample comments from high response subcategories.